Universität Hamburg

DER FORSCHUNG | DER LEHRE | DER BILDUNG

Super-Natural Interaction: An Enactivism-Based Conceptualization and Applications in Immersive Telementoring

Dissertation

with the aim of achieving a doctoral degree Dr. rer. nat.

at the Faculty of Mathematics, Informatics and Natural Sciences

Department of Informatics

Universität Hamburg

Bastian Dewitz

2025

Dissertation Committee:

First reviewer:	Prof. Dr. Frank Steinicke
Second reviewer:	Prof. Dr. Gerd Bruder
Third reviewer:	Prof. Dr. Christian Geiger
Chair of the examination commission:	Prof. Dr. Eva Bittner
Deputy chair of the examination commission:	Prof. Dr. Matthias Rarey

Date of Disputation: 23.04.2025

ACKNOWLEDGMENTS

First, I would like to thank my supervisors, Prof. Dr. Frank Steinicke and Prof. Dr. Christian Geiger, for giving me the opportunity to work on this scientific topic. I am deeply grateful for their support and guidance during this endeavor, as well as for the freedom I was given in approaching this thesis – especially considering the difficult time of working on this dissertation during the COVID-19 pandemic. I am also very thankful to Prof. Dr. Gerd Bruder for additional support in the final phase of this dissertation and for reviewing this thesis. Furthermore, I would like to thank the chairs of the examination commission, Prof. Dr. Eva Bittner and Prof. Dr. Matthias Rarey, for conducting the formal part of the evaluation of this thesis, as well as Anna Leffler for her support throughout the process.

I also want to thank Dr. Falko Schmid, PD Dr. Hug Aubin, and Prof. Dr. Artur Lichtenberg for providing the amazing opportunity to conduct the practical part of my research in the cardiac clinic of the university hospital Düsseldorf. Being allowed to evaluate my research topic within the real-world context of heart transplantation was an extraordinary experience. Again, I highly appreciate that I was entrusted to approach my research topic very freely, which immensely supported this thesis. Although the work in the medical domain, with its requirements, responsibilities, and constraints on research, made things much more complicated, it also made this research very interesting and meaningful.

I thank my co-authors, colleagues, and students at MIREVI lab at Hochschule Düsseldorf, HCI group at the University of Hamburg, and DHLD at the University Hospital of Düsseldorf, who directly or indirectly supported this dissertation. Many colleagues have become close friends, and I am very grateful for all the great memories we have made during this journey.

Last but not least, I am grateful to my friends and family, who, in their own unique ways, have supported and encouraged me throughout this journey. Without you having my back, finishing this thesis would not have been possible!

ABSTRACT

Immersive virtual environments (IVEs) enable the implementation of interaction techniques that are not feasible in the physical world. For a virtual avatar in a virtual world that is not limited to the laws of the real world, it is a simple task to fly, teleport, or interact with remote objects through telekinesis. This type of interaction has been called "super-natural" interaction. However, a precise definition of "super-natural" is still lacking in current research, and the concept is usually only described on a technical level with a focus on objective qualities and without considering the induced subjective effects on the user. Moreover, other terms exist that, on a conceptual level, are used to describe similar ideas, such as "magic," "non-isomorphism," or "superpowers." Currently, these terms are used interchangeably and without clear delineation.

In this dissertation, the nature and implications of super-natural interaction are explored to further characterize this concept, with the goal of clarifying what makes this approach to interaction design beneficial and opening the scientific discourse toward subjective and phenomenological effects. The thesis is divided into two parts; the first presents an enactivism-based conceptualization, and the second presents applications of super-natural interaction in immersive telementoring.

In the **first part** of this thesis, concepts that describe the constituents of "super-natural" interaction are developed from an enactivism-inspired perspective, including related fields of philosophy and cognitive science, such as post-phenomenology, pattern theory of self, and schema theory. These concepts are combined in a conceptual framework that allows for i) the classification of interaction techniques based on the proposed properties internalizability, congruence, and enhancement, and ii) the delineation of diverse types of interaction techniques from one another. Overall, the presented conceptualization provides a framework for analyzing human-computer interaction on a subjective and phenomenological level, with a strong focus on embodiment, which complements the objectivist perspective typically considered in the field of human-computer interaction.

In the **second part**, the application of super-natural interaction techniques is analyzed in the context of heart transplantation as an important real-world application. Super-natural interaction techniques are often implemented in experimental setups or in the context of gaming. Applying these concepts to support real-life activities, however, can be challenging. Especially in the medical domain, surgeons and health professionals remain reluctant to use virtual reality (VR) technology. To systematically describe the challenges of introducing VR technology as a transformative technological intervention, we present the ESTA framework and apply it in the context of heart transplantation. Super-natural interaction techniques seem promising in enabling the efficient performance of tasks without a steep learning curve, which mitigates difficulties in handling VR technology in the medical domain. As a first step in analyzing the use of super-natural interaction techniques, a mobile prototype system for immersive telementoring has been developed that provides a platform for novel interaction techniques. As a second step, novel super-natural interaction techniques tailored for exploration and annotation tasks in immersive telementoring are presented and evaluated to determine if this approach is reasonable in this context.

ZUSAMMENFASSUNG

Immersive virtuelle Umgebungen ermöglichen die Implementierung von Interaktionstechniken, die nicht in der Realität genutzt werden können. Für einen virtuellen Avatar, der nicht durch die Gesetzte der realen Welt eingeschränkt wird, ist es einfach, zu fliegen, sich zu teleportieren oder Objekte aus der Ferne durch Telekinese zu manipulieren. Diese Form der nichtrealistischen Interaktion wird unter anderem als "super-natürlich" (engl. "super-natural") bezeichnet. Die Nutzung dieses Begriffs in der Forschung is jedoch nicht unproblematisch, da es keine präzise Definition für diesen Begriff gibt und das Konzept häufig nur auf einer technischen Ebene mit einem Fokus auf objektive Faktoren beschrieben wird, ohne die subjektive Effekte auf Nutzer zu betrachten. Zusätzlich existieren weitere Begriffe, die, auf einer konzeptionellen Ebene, genutzt werden, um ähnliche Ideen zu beschreiben, beispielsweise "magisch", "nicht isomorph", oder "Superkräfte". Eine klare Abgrenzung wird nicht vorgenommen und die Nutzung erfolgt scheinbar willkürlich und austauschbar.

In dieser Dissertation werden die Eigenschaften und Implikationen von super-natürlicher Interaktion untersucht, um das Konzept tiefergehend zu beschreiben. Das Ziel ist, herauszuarbeiten, was diesen Ansatz für Interaktionsdesign vorteilhaft macht und den wissenschaftlichen Diskurs hin zu subjektiven und phänomenologischen Effekten zu öffnen. Die Thesis ist in zwei Teile geteilt; der erste präsentiert eine auf Enaktivismus basierende Konzeptualisierung, der zweite die Anwendung von super-natürlicher Interaktion im Kontext von immersivem Telementoring.

Im **ersten Teil** der Thesis werden Konzepte entwickelt, die von Enaktivismus und verwandten Forschungsbereichen wie Post-Phänomenologie, Pattern Theory of Self und Schematheorie abgeleitet werden, um die Kernaspekte von super-natürlicher Interaktion herauszuarbeiten. Verschiedene Konzepte werden in ein konzeptionelles Framework zusammengefügt, das i) die Klassifikation von Interaktionstechniken anhand von den vorgeschlagenen Eigenschaften Internalisierbarkeit (Internalizability), Kongruenz (Congruence) und Verbesserung (Enhancement) erlaubt und ii) die Abgrenzung der Begriffe untereinander ermöglicht. Insgesamt bietet die vorgestellte Konzeptualisierung ein Framework, das genutzt werden kann, um Mensch-Computer-Interaktion auf einer subjektiven und phänomenologischen Ebene mit einem starken Fokus auf körperliche Aspekte zu untersuchen, was die objektivistische Perspektive, die typischerweise in der Forschung zu Mensch-Computer-Interaktion eingenommen wird, komplementiert.

Im **zweiten Teil** wird die Anwendung super-natürlicher Interaktion im Kontext von Herztransplantationen als wichtiger Anwendungsfall in der realen Welt untersucht. Supernatürliche Interaktionstechniken werden häufig in experimentellen Systemen oder im Kontext von Gaming eingebunden. Eine Nutzung zur Unterstützung von Aktivitäten in der realen Welt kann hingegen herausfordernd sein. Insbesondere in der medizinischen Anwendung sind Operateure und Experten oft nicht von der Anwendung von VR-Technologie überzeugt. Um die Herausforderungen bei der Einführung von VR Technologie systematisch zu beschreiben, wird das ESTA Framework präsentiert. Zur Untersuchung supernatürlicher Interaktionstechniken, wurde als erster Schritt ein mobiler Prototyp entwickelt, der als Plattform zur Implementierung super-natürlicher Interaktionstechniken dient. Im zweiten Schritt werden neuartige super-natürliche Interaktionstechniken, die auf die Anforderungen im immersiven Telementoring zugeschnitten sind, vorgestellt und evaluiert, um zu überprüfen, ob sie in diesem Kontext einen sinnvollen Ansatz darstellen.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PART I

FOUNDATIONS

CHAPTER 1

INTRODUCTION

1.1 Motivation

The use of technology is a fundamental aspect of human life that allows us to overcome physical and cognitive limitations. For millennia, technology and tools profoundly impacted our society and shaped the ways in which we interact with the world. With digital technology, human life and culture have undergone a rapid and transformative evolution that far surpasses the rate of pre-industrial technology. Particularly in the field of human-computer interaction (HCI), the ways in which humans and computer systems interact have changed significantly. Early computation machines, such as the ENIAC, Z3, and MK2, were operated using artificially created abstract languages, and interaction with these machines was not real-time but instead followed a sequential process of symbolic input and output. This initially highly artificial and abstract interface layer between humans and computers has steadily evolved over the past decades, and these advances in technology have led to more intuitive and natural ways for users to engage with computer systems. Nowadays, computer interfaces are often seamlessly integrated into our everyday lives, and complex systems can be controlled using intuitive means of input. The development of user-friendly interfaces has driven the widespread adoption of computer technology, and subsequently, the number of users has increased drastically. Still today, the ongoing research and development activities produce new devices and technology-based interactions that find their way into our lifeworld. One approach in HCI that has gained major attention for its unique possibilities of interaction over the past decade is VR.

VR systems have been imagined in science fiction literature long before an actual implementation was possible. A fictional work that shows many parallels to our idea of virtual reality is the short story *Pygmalion's Spectacles* [Wei16] by Stanley G. Weinbaum, published in 1935. In this short story, the protagonist uses a fictitious pair of magic spectacles that reproduce sensations of sight, sound, smell, taste, and touch. These glasses are capable of fully simulating artificial characters, so-called "shadows," that interact with the protagonist and seem like reality. The 1950s novel *The Veldt* by Ray Bradbury presents an immersive room that displays fictitious locations for entertainment, resembling today's CAVE systems [CN+92]. An influential essay in the context of HCI is Ivan E. Sutherland's *The Ultimate Display* [Sut65], published in 1965, in which a fictional artificial environment is described that is indistinguishable from reality, which has served as an inspiration for VR for many researchers. The concept of immersive reality simulations was further popularized in fictional works such as the *Holodeck* in *Star Trek* or the movie trilogy *The Matrix*. The first functional VR prototype systems were developed in the 1960s using, from today's perspective, much simpler technology. The limitations imposed by the available computer hardware made the practical application of these systems difficult, and it took several decades to make VR technology available for everyday use [Kus14].

In 2016, the "year of virtual reality" [Ste16a], significant steps were made to bring this technology from research laboratories to the real world. Important consumer products, such as HTC's Vive System and Oculus Rift CV were released, which made this technology both affordable and usable. The potential of VR technology has also been acknowledged

outside of the research community. According to the Gartner Hype Cycle for Emerging Technologies, VR technology reached its *Plateau of Productivity* in 2018, which, in terms of the Gartner Hype Cycle, means that VR has matured and its full potential can be harvested for business purposes[1]. In recent years, large technology companies have claimed terms in the context of VR to promote their products and visions of variants of virtual reality and related technologies. In 2016, Microsoft introduced 'Windows Mixed Reality,' which "blends people, places, and objects to create exciting new experiences across the physical and virtual worlds" [Mic16]. In Meta's vision from 2021, "the next platform and medium [after desktop PCs and smartphones] will be even more immersive; an embodied internet where you are in the experience, not just looking at it. And we call this the Metaverse" [Met24]. Recently, in 2024, with the introduction of Apple's Vision Pro headset, Apple proclaimed: "The era of spatial computing is here, where digital content blends seamlessly with your physical space" [App23]. With these advances, interaction with realistic virtual content, or a mixture of physical space and digital data, may become a usual experience in the near future. Diverse fields of work can potentially benefit from this technology, and everyday interaction may change in a way similar to the transformation that society and everyday life have undergone with the ubiquitous availability of smartphones nowadays. A key component in this endeavor is the development of interaction paradigms that are capable of providing both efficient and natural-feeling ways of interaction within the novel virtual worlds.

In his visionary essay, *the Ultimate Display*, Sutherland imagined hypothetical virtual reality systems with an ability to perfectly simulate a physical environment in which interaction faithfully recreates reality:

> "The ultimate display would, of course, be a room within which the computer can control the existence of matter. A chair displayed in such a room would be good enough to sit in. Handcuffs displayed in such a room would be confining, and a bullet displayed in such a room would be fatal." [Sut65]

In such a system, humans would be able to interact freely and naturally with virtual content in the same way as they interact with the real world. The virtual world would be a perfect copy of the real world. However, this is not where the application of VR stops; he further imagined VR systems that not only offer the possibility of creating environments that mimic reality to provide natural ways of interaction but also enable the implementation of novel forms of interaction that exceed what is possible in the real world, such as flying, x-ray vision, or teleportation.

> "There is no reason why the objects displayed by a computer have to follow the ordinary rules of physical reality with which we are familiar. The kinesthetic display might be used to simulate the motions of a negative mass. The user of one of today's visual displays can easily make solid objects transparent - he can "see through matter!" ... With appropriate programming such a display could literally be the Wonderland into which Alice walked." [Sut65]

A term in research that has been used to describe interaction in VR that is not limited to the rules of reality is "super-natural." These "super-natural" techniques have been researched for decades, and their use is well-accepted within the VR research community. However, the terminology is not well-defined, and various other terms exist that describe similar interaction techniques. Researchers not only use the term "super-natural", but also, for example, "superpowers", "hyper-natural", "magic", and "empowerment", often without further distinguishing these concepts or providing a precise definition (see chapter

[1] The Gartner Hype Cycle presents a business perspective on technology [SL10] and started featuring VR technology in 2004. Since 2018, VR has not been featured in the Gartner Hype Cycle anymore.

3). Currently, the use of terms seems more related to personal taste than explicit scientific definitions, and often, the focus of research lies on purely objective criteria, such as input fidelity [MLP16] or task performance. Therefore, the first part of this thesis aims to provide i) a concise terminology and conceptual framework as a foundation for the delineation of terms and ii) an analysis of the subjective and phenomenological effects of implementing VR interaction that purposefully rejects reality.

At the beginning of this dissertation project, the first naïve approach for classification, based on the experience of interaction in VR and common patterns of design, was to analyze interaction techniques concerning their 'realism,' 'empowerment,' and 'naturalness.' [Dew+18]. However, grasping these concepts fully and answering entailing questions, such as, "What is reality?" or "What is natural to humans?", is hardly possible and depends on the scientific and philosophical framework. In this thesis, the enactive approach is chosen as an interesting contemporary emerging theoretical framework in cognitive science that challenges the traditional view of human cognition as an information-processing system. With its perspective on subjective experience, it provides new perspectives on interaction and offers new insights into the lived experience of acting within a non-reality-based environment. The enactive approach is further combined with other philosophical concepts to form a coherent framework for describing technology-mediated human-world relations with a special focus on the effects of VR.

Furthermore, the use of these "super-natural" interaction techniques is investigated in the context of immersive telementoring as a domain of practical real-world application. Immersive telementoring refers to the practice of remotely supporting novice surgeons, the so-called *mentees*, by expert surgeons from afar, the *mentors*, using immersive technology. With recent technology, such as 3D depth sensors, real-time video transmission, and VR, such systems can be designed to allow expert surgeons to experience a remote surgical procedure as if they were standing in the operating room. This can potentially increase the surgery outcomes in comparison to traditional methods such as phone calls or 2D video transmission. However, in the context of immersive telementoring, the topic of providing engaging and efficient interaction techniques can be considered underexplored. Especially the use of VR technology within the conditions of real-world hospital environments with their specific requirements has not often been investigated. Both aspects are addressed in the second part of this thesis, in which an immersive telementoring system using contemporary hardware is developed for the use case of heart transplantation, and several "super-natural" interaction techniques tailored for this task are implemented.

The conceptual framework in the first part of this thesis and the practical application in the second part were carried out in parallel, and both have influenced each other (see Fig. 1.1). This allowed practical findings to influence the theoretical work, and conceptual ideas to be analyzed within a practical application.

1.2 Research Questions

The first part encompasses the conceptual aspects of analyzing super-natural interaction. The ontological questions researched are: "What makes an interaction technique super-natural?" and "What are the subjective effects of using "super-natural interaction?" Three sub-questions (R1, R2, R3) are formulated in the first part of this thesis. The second part of this thesis focuses on the practical application of super-natural interaction under real-world conditions. To analyze how super-natural interaction in VR and immersive technology can be implemented in real-world scenarios, two further questions are formulated (R4, R5).

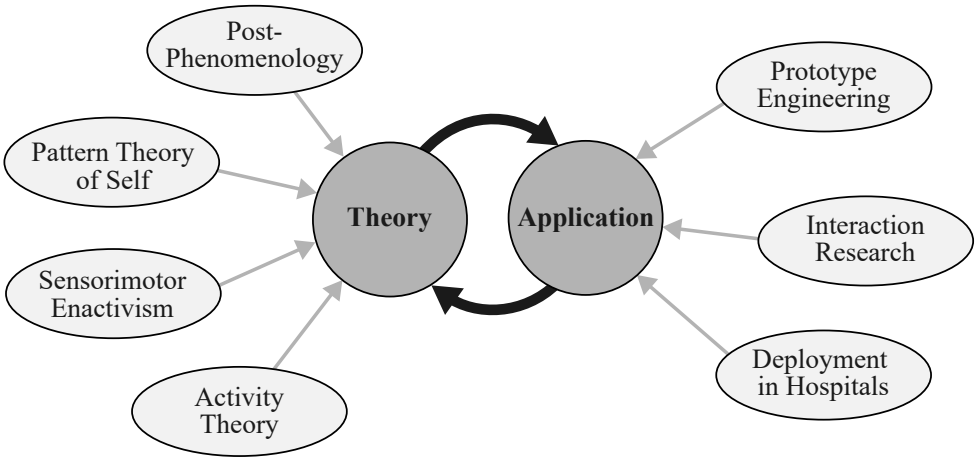


Figure 1.1: Theoretical influences and important aspects of the practical application that are researched in this thesis.

R1: How is super-natural interaction described in research literature?
The term 'super-natural' has been used in several research publications regarding interaction in VR. This question aims at clarifying the central and recurring themes that are mentioned in the context of super-natural interaction by researchers.

R2: How can super-natural interaction be described within the framework of enactivism?
Enactivism is a theoretical framework for cognition that has not often been used to describe aspects of VR interaction. It provides a subjective and phenomenological perspective on interaction and cognition that complements the objective findings presented in the literature.

R3: How is super-natural interaction related to other types of interaction?
Many different terms are used in research to describe interaction techniques, however, in many cases, without an explicit definition. 'Super-natural' interaction techniques are assumed to constitute a distinct class of interaction techniques within the design space of interaction techniques, characterized by specific properties and relationships to other classes.

R4: How can heart transplantation be supported using super-natural interaction and state-of-the-art immersive technology?
The use of immersive technology is not common in hospital environments, and immersive systems for heart transplantation do not yet exist. Transforming existing procedures with state-of-the-art technology presents unique challenges to research and development processes that need to be addressed.

R5: How can super-natural VR interaction be applied in the context of exploration and annotation in immersive telementoring?
Annotation and exploration are fundamental tasks required for immersive telementoring, which can be addressed using super-natural interaction techniques tailored to the use case of heart transplantation. The goal of these super-natural techniques is to be both efficient and natural to use.

1.3 Thesis Outline

This thesis is divided into four parts:

Part I introduces this research work. **Chapter 1** provides an overview of the motivation, structure, and contribution of this thesis. **Chapter 2** presents the scientific foundations of this thesis in the areas of human-computer interaction, virtual reality, and human factors that are relevant in the subsequent conceptual and practical parts.

Part II encompasses the conceptual part of this research. **Chapter 3** begins with a literature review on 'super-natural' interaction in VR, from which a working definition for super-natural interaction techniques with two properties is derived: They i) reduce realism in order to provide enhanced functionality, and they ii) are easy to learn and use. In **Chapter 4**, the enactive approach and related philosophical concepts are applied to analyze the first property (reduced realism) and its effects on our perception of ourselves and the world. Finally, the concept of the *technological pattern of self* and the derived terms *congruence* and *enhancement* are presented (**R2**). **Chapter 5** analyses the second property (easy to learn and use) and proposes the property *internalizability* based on the concept of interaction schemata (**R2**). **Chapter 6** presents the 'ICE cube' as a conceptual framework for identifying classes of interaction techniques based on their internalizability, congruence, and enhancement (**R3**). To validate this framework, a pilot questionnaire is constructed and analyzed in an exploratory factor analysis, aiming at a preliminary validation of the proposed concepts.

Part III describes the practical application of VR technology and super-natural interaction in immersive telementoring. **Chapter 7** explains the use case and the derived 'ESTA framework' as a general model for technological interventions. The ESTA framework is further used to analyze the intended prototype system and define the challenges that are addressed in this thesis (**R4**). **Chapter 8** illustrates the technical challenges of the developed prototype system, including an evaluation of the prototype system with medical experts that demonstrated the feasibility of this approach. **Chapter 9** demonstrates the use of super-natural interaction techniques for object exploration (**R4, R5**), and in **Chapter 10**, the use of super-natural interaction techniques for surface annotations is analyzed (**R4, R5**).

Part IV is the final part of this research work. In **Chapter 11**, the research findings, as well as limitations and potential for future work, are discussed. In the final **Chapter 12**, a conclusion for this thesis is drawn.

1.4 Contribution

First, this thesis aims at clarifying the terminology used to classify interaction techniques, specifically the term 'super-natural.' The approach relies on the subjective analysis and interpretation of aspects of phenomenology, embodied interaction, and enactivism in the context of VR interaction. The following contributions are made:

- A literature review on super-natural interaction in VR.
- The phenomenology-based concepts of the 'technological pattern of self' (derived from pattern theory of self [Gal13]) and 'congruence' that incorporate different philosophical concepts.
- The concept of 'interaction schemata' and 'internalizability' in which sensorimotor enactivism, image scheme theory, and embodied interaction are combined.
- A framework for classifying interaction techniques in VR, the 'ICE cube.'

Second, the benefits of immersive technology and super-natural interaction are analyzed in the domain of heart transplantation as both a challenging and important real-world application. This part includes the following contributions:

- An evaluated prototype system for immersive telementoring in heart transplantation using state-of-the-art technology that has been tested under real-world conditions.
- A general framework for analyzing challenges of deploying novel technology in established domains of work, the 'ESTA-framework'.
- Four novel interaction techniques for annotation in the context of immersive telementoring: Aperture Zoom, Portal Zoom, LensDraw, PalmDraw.

1.5 Publications

Main authorship As the primary author, the following peer-reviewed publications have been published in the course of this dissertation. Publications that are fully included in this dissertation and adaptations are marked with a '▶', whereas partially included publications are marked with a '▷'. Appendix A presents a summary of the work contribution for each publication.

2023 ▶ *Bastian Dewitz, Christian Geiger, Frank Steinicke.* Enacting Interaction in Virtual Reality. In GI VR/AR Workshop. Gesellschaft für Informatik eV.

2023 ▶ *Bastian Dewitz, Sobhan Moazemi, Sebastian Kalkhoff, Steven Kessler, Christian Geiger, Frank Steinicke, Hug Aubin, and Falko Schmid.* Enacted Selves in Technological Activities – Framework and Case Study in Immersive Telementoring. In Proceedings of Mensch und Computer 2023.

2023 ▶ *Bastian Dewitz, Sukran Karaosmanoglu, Robert W. Lindemann, Frank Steinicke.* Magic, Superpowers, or Empowerment? A Conceptual Framework for Magic Interaction Techniques. In IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW) (pp. 807-808). IEEE. (Poster)

2022 ▶ *Bastian Dewitz, Christian Geiger, Frank Steinicke.* Acting Beyond Reality – The Role of Schemata in Mixed-Reality Super-Natural Interaction. In GI VR/AR Workshop. Gesellschaft für Informatik eV.

2022 ▶ *Bastian Dewitz, Roman Bibo, Sebastian Kalkhoff, Sobhan Moazemi, Artur Liebrecht, Christian Geiger, Frank Steinicke, Hug Aubin, Falko Schmid.* Towards 5G Telementoring in VR-Assisted Heart Transplantation Using HoloLens 2. In GI VR/AR Workshop. Gesellschaft für Informatik eV.

2022 ▶ *Bastian Dewitz, Roman Bibo, Sobhan Moazemi, Sebastian Kalkhoff, Stephan Recker, Artur Lichtenberg, Christian Geiger, Frank Steinicke, Hug Aubin, Falko Schmid.*Real-Time 3D Scans of Cardiac Surgery Using a Single Optical-See-Through Head-Mounted Display in a Mobile Setup. In Frontiers in Virtual Reality, 137.

2021 ▶ *Bastian Dewitz, Calvin Huhn, Christian Geiger, Frank Steinicke.* Virtuality between my Fingers – Investigation of Zoom Mechanisms for Visual Exploration of Virtual Environments. In GI VR/AR Workshop. Gesellschaft für Informatik eV.

2021 *Bastian Dewitz, Christian Geiger, Frank Steinicke.* Virtual Visus – Vision Acuity and Text Legibility in Virtual Environments. In GI VR/AR Workshop. Gesellschaft für Informatik eV.

2020 ▷ *Bastian Dewitz, Christian Geiger, Frank Steinicke.* Hand-Based Interaction on a Millimeter Scale in Virtual and Augmented Reality. In GI VR/AR Workshop. Gesellschaft für Informatik eV. (Poster)

2019 ▷ *Bastian Dewitz, Frank Steinicke, Christian Geiger.* Functional Workspace for One-Handed Tap and Swipe Microgestures. In Mensch und Computer 2019-Workshopband. VARECo workshop.

2018 ▷ *Bastian Dewitz, Philipp Ladwig, Frank Steinicke, Christian Geiger.* Classification of Beyond-Reality Interaction Techniques in Spatial Human-Computer Interaction. In Proceedings of the Symposium on Spatial User Interaction (pp. 185-185). (Poster)

Co-authorship Contributions that influenced this dissertation or that were influenced by this dissertation project were made to following publications as co-author:

2022 *Artur Liebrecht, Roman Bibo, Bastian Dewitz, Sebastian Kalkhoff, Sobhan Moazemi, Markus Rollinger, Jean-Michel Asfour, Klaus-Jürgen Janik, Artur Lichtenberg, Hug Aubin, Falko Schmid.* ARMAGNI: Augmented Reality Enhanced Surgical Magnifying Glasses. In Scandinavian Conference on Health Informatics (pp. 46-51).

2021 *Charlotte Triebus, Ivana Družetić, Bastian Dewitz, Calvin Huhn, Paul Kretschel, Christian Geiger.* is a rose – A Performative Installation between the Tangible and the Digital. In Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction (pp. 1-4).

2021 *Tobias Picker, Bastian Dewitz, Christian Geiger, Frank Steinicke.* Echtzeit-Fingertracking in Unity 3D durch 3D Convolutional Neural Networks in einem Multi-Depth-Camera-Setup. (engl. 'Real-Time Finger Tracking in Unity 3D using 3D Convolutional Neural Networks in a Multi-Camera Setup.') In GI VR/AR Workshop. Gesellschaft für Informatik eV.

2020 *Charlotte Triebus, Bastian Dewitz, Ivana Družetić, Calvin Huhn, Paul Kretschel, Christian Geiger.* is a rose – A Performative Installation in the Context of Art and Technology. In: Kultur und Informatik: Extended Reality. 2020, pp. 153–165.

2019 *Philipp Ladwig, Bastian Dewitz, Hendrik Preu, Mitja Säger.* Remote Guidance for Machine Maintenance Supported by Physical LEDs and Virtual Reality. In Proceedings of Mensch und Computer 2019 (pp. 255-262).

2018 *Marcel Tiator, Christian Geiger, Bastian Dewitz, Ben Fischer, Laurin Gerhardt, David Nowottnik, Hendrik Preu.* Venga! Climbing in Mixed Reality. In Proceedings of the First Superhuman Sports Design Challenge: First International Symposium on Amplifying Capabilities and Competing in Mixed Realities (pp. 1-8).

2018 *Tobias Picker, Bastian Dewitz, Christian Geiger.* Fingertracking durch neuronale Netze anhand reduzierter Markersets und Motion-Capture-Daten. (engl. 'Finger Tracking Utilizing Reduced Marker Sets and Motion Capture Data.') In GI VR/AR Workshop. Gesellschaft für Informatik eV.

2018 *Alexander Giesbrecht, Sarah von Styp-Rekowski, Bastian Dewitz, Christian Geiger.* Examining effects of altered gravity direction in Room-Scale VR. In GI VR/AR Workshop. Gesellschaft für Informatik eV.

CHAPTER 2

INTERACTIVE VIRTUAL WORLDS

2.1 Human-Computer Interaction

Human-computer interaction can be defined as "a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them" [Hew+92, p. 5]. It analyzes the nature of HCI systems, the characteristics of human users, the properties of computers and their use, and the development process of technology utilizing knowledge and methods from diverse disciplines, such as cognitive psychology, sociology, computer science, and design. This section provides a brief overview of important aspects in HCI research.

2.1.1 The User Interface

The term *interface* is a neologism introduced by James Thomas in 1869 to describe the physical phenomenon of a dividing surface between two expanding fluids [Tho16]. The term has been adapted in HCI to describe a mediating layer between the user and the computer system which separates the internal mechanisms of both and translates messages between the user language and the system language [Dix+04] to couple an *explicit-human process* to an *explicit-artifact process*, in this case, a program running on a computer system [Eng23]. The interface is the "medium through which the communication between users and computers takes place" [LJ+17, p. 6], which conceals the complex mechanisms of computer systems behind an additional layer that focuses on the human user, rather than the binary computer language. The properties of the mediating interface layer depend on various factors, e.g., the available technology, context of use, and goals of interaction. Depending on the approach and focus of research, various models have been proposed to describe the interaction between users and computer systems. Over time, these interface models shifted their focus from the computer's technical requirements and functionality to the user's perspective on interaction and the psychological factors involved.

2.1.2 Interfaces and Models

HCI and research regarding user interfaces emerged from the difficulties of controlling early computer systems. These systems were machine-oriented, and users had to adopt the binary computer language to successfully interact with computers. The direct interaction with computer systems without a proper interface layer was challenging, and "it would [have been] difficult to design a language that was more difficult or unnatural for human beings to learn to use" [Mil17]. The instruction of computer systems was therefore facilitated by the creation of "arbitrary and precise command languages" [WW11, p. 38] that were inspired by human language and human ways of thinking [Lin66]. Grace Hopper, who was a leading scientist in the development of the first compilers, stated in 1980 that the goal of her work on programming languages in the 1950s was to make computer systems "... easier to use by everybody. Feeling that we needed greater use of it, more information processing, something had to be done to get people to use it" [Pan80]. The development of compilers paved the way for Command Line Interfaces (CLI)s and the widespread use of computer systems by "[bringing] the interface between machine and [humans] a giant step

closer to the [humans], and in the same step ... whole new classes of users into interaction with computers" [Lin66]. A central concept of these interfaces is the dialogue between a human operator and a system in which messages are exchanged following an established convention [Nie86] in the form of semantic tokens in a precisely defined system syntax that correspond to a system-specific execution of computer code [Bux83]. This mode of interaction was functional, but improvement regarding user-friendliness was still possible. Sutherland wrote in 1964 that the human operators of computer systems "have been writing letters to, rather than conferring with ... computers. For many types of communication ..., typed statements can prove cumbersome" [Sut64]. This difficulty in handling computer systems limited the potential of computer systems, which was also expressed by the Air Force Cambridge Research Laboratories in 1967: "Computers are capable of many more tasks than they are presently assigned, primarily because no one knows how to go about instructing the computer on how to do more difficult tasks" [Lab67, p. 13].

In the 1960s, experimental systems introduced novel ways of interaction, such as the first touch-sensitive display [Joh65], graphical user interfaces [Sut64], and also the first head-mounted display [Sut68]. In the so-called "Mother of all demos" on December 9th, 1968, Engelbart presented, among other influential advanced concepts in HCI, the *mouse* as a novel input device [Int18; EE68] to the public, which, still today, remains an integral part of many modern computer systems. The new emerging style of interaction with computer systems in the 1980s was the Graphical User Interface (GUI), including the WIMP (windows, icons, menus, pointers) concept, in which the user interacts using metaphorical representations of actions. The Xerox Alto was the first commercially available computer system to consistently follow the WIMP concept and can, therefore, be considered the common ancestor of today's personal computers, even though only a limited number of these systems were deployed. In 1984, the Macintosh was released by Apple, which claims the title of the first successful personal computer system [Nor98]. Today, CLIs and GUIs are the dominating concepts of how users interact with computer systems, including personal computers, web-based interactions, and mobile devices.

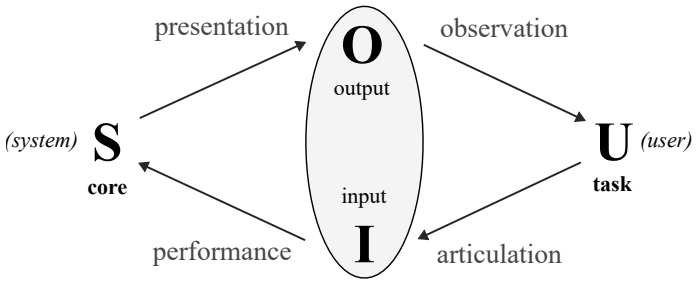


Figure 2.1: Abowd and Bale's general interaction framework describes the interface as a mediator between the user and a system. Adapted from [AB91].

The interaction between a user and a computer system can be described from various perspectives through frameworks and models. The interpretation of processes, as well as the produced insights into the interaction cycle, are dependent on the theoretical assumptions. In Abowd and Bale's interaction framework [AB91] (see Fig. 2.1), the focus lies on the general exchange of information within the user-computer system, using the interface as a translator between different languages. In this model, a user formulates a task and performs actions directed towards the interface (*articulation*). The interface *translates* the user's input language to the system's core language (*performance*), and the system reacts to the input by executing some binary machine code. This changes the system's state and provides an output to the interface (*presentation*). Finally, the interface translates the new system state to the user's language, which is perceived (*observation*).

In CLIs, the interface layer corresponds to the translation of language-like typed commands (articulated by the user) to machine code as system input. Conversely, the presentation of the results of the code execution is translated into readable text as output. In GUIs, on the other hand, the performed actions are less precisely defined and more related to physical interaction with objects in the real world. A key component of GUIs is visibility, which enables users to explore an application and facilitates visual learning and memory [Nor10]. Another key component is the direct manipulation of digital content that resembles the physical manipulation of objects in the real world, a concept first introduced in Sketchpad [Mac12]. Shneiderman formulated three key properties of direct manipulation: i) a continuous representation and visibility of the objects (and data) and user actions, ii) physical interactions with objects as user input, and iii) incremental and reversible actions with an immediate effect on objects [Shn97]. In traditional 2D GUI interfaces, six types of interaction techniques are commonly implemented following certain conventions to allow users to use different computer systems without having to learn new forms of interaction: *select*, *position*, *orient*, *path*, *quantify*, and *text* [FWC84].

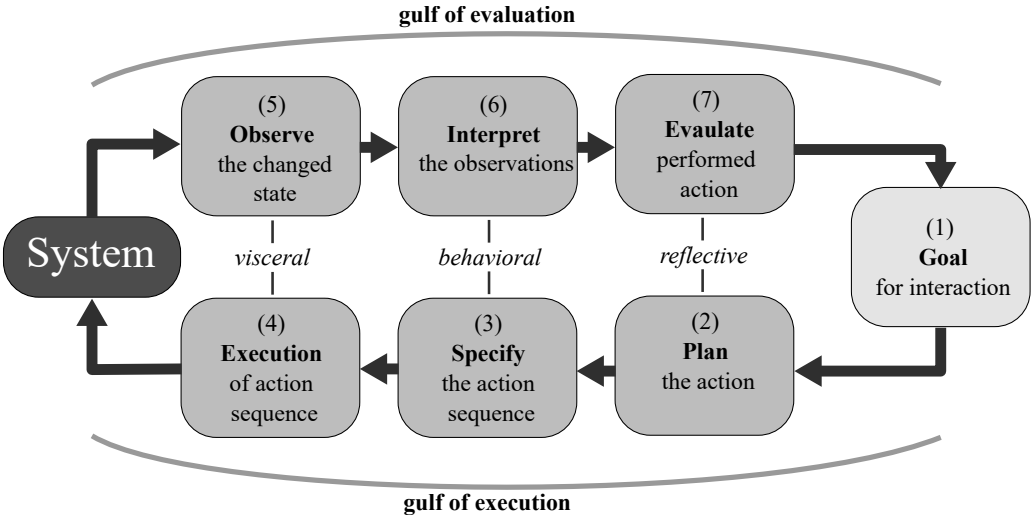


Figure 2.2: Norman's model of the seven stages of action with the gulfs of evaluation and execution, which emphasize the human operator's perspective in HCI. Adapted from [Nor13].

A highly influential model in HCI with a strong emphasis on the user's perspective on the interaction is the *seven stages of action* model (see Fig. 2.2) introduced by Don Norman, which describes distinct steps that have to be executed by a user during the interaction with a system [Nor13]: First, a goal is formulated (1) that describes what has to be achieved by the interaction. From this goal, the user derives the intention to act and plans how to interact with the system specifically (2). This way of interaction typically involves a sequence of actions (3) that are required to be performed to achieve the goal. This sequence is executed by physical motor actions (4), such as typing on a keyboard, using voice, or pressing buttons on a controller. The physical action produces the input for the system, which performs actions that are not visible to the user and produces some form of perceivable output. The changed state of the system is perceived (5) by the user with one or more of the available sensory channels, for example, visual feedback, auditory cues, or tactile vibrations. The provided perceptions need to be interpreted (6), which allows for a complete evaluation of whether the initially formulated intention of achieving a goal by the chosen means was successful (7). Depending on the evaluation, intrapersonal, and external factors, the user formulates a new goal for interaction. In Norman's model, three layers of action are distinguished. The highest *reflective* layer is determined by conscious thought

and active thinking. On the *behavioral* layer, acting is characterized by the unconscious execution of learned skills and automatic responses. The lowest *visceral* layer describes the most fundamental level of processing that encompasses basic primordial responses, e.g., fight or flight responses, with an immediate and completely unconscious coupling to the physical body. The model can be further divided into a *gulf of execution* that encompasses the progression from the user's initial goal to performing motor actions (phases 1 to 4, corresponding to the *articulation*), and a *gulf of evaluation*, which describes the transition from perceiving the system's output to evaluating the effect of actions (phases 5 to 7, corresponding to the *observation*). Visually speaking, these gulfs need to be overcome by users to enable them to anticipate the outcome of actions to interact successfully with the system.

2.1.3 Interaction Techniques and Metaphors

Interfaces typically implement several interaction techniques that are aimed at solving specific tasks within the system using the provided means of the interface. Interaction techniques can be defined as the "fusion of input and output, consisting of all software and hardware elements, that provides a way for the user to accomplish a task" [Tuc04]. They are interface-dependent and provide one distinct way to perform a specific task. Often, techniques are hierarchically structured (see Fig. 2.3) and involve multiple subtasks. For each subtask, multiple *technique components* may exist as distinct ways of performing a subtask. An interaction technique can be described by all technique components that are selected and implemented by the designer or developer [BH99].

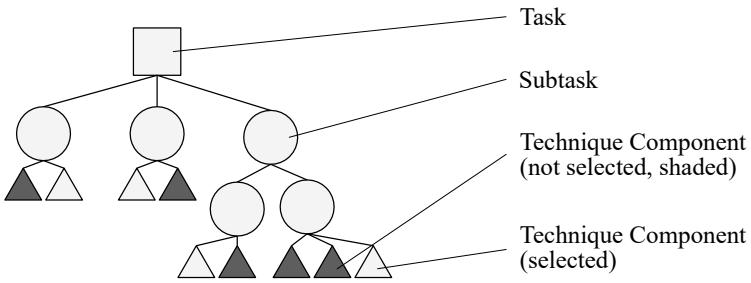


Figure 2.3: The hierarchical structure of tasks, subtasks, and technique components. Adapted from [BH99].

To allow easier access to user interfaces and system functionality, especially in the case of novice users, interaction techniques often rely on the idea of implementing easily recognizable, understandable, and memorable metaphors [Mar93]. Metaphors are used "to increase the initial familiarity of actions, procedures and concepts by making them similar to actions, procedures and concepts that are already known" [CMK88, p. 67] to improve the ease of learning and appeal to users [Mar98]. Including metaphors into a system "extends the intentional range ... by providing new ways to conceive of one's actions in the system, and providing new entities [for interaction]" [Dou04, p. 143]. A taxonomy that is heavily influenced by image scheme theory (see section 2.3.1) is provided by Barr et al. [BBN02], in which four types of interface metaphors are distinguished:

- *Orientational metaphors* transform concepts of 2D or 3D space into input interface concepts, including physical concepts, such as vertical and horizontal sliders, but also abstractions, such as associating a happy mood with 'up.'
- *Ontological metaphors* explain interaction in terms of existing in a physical three-dimensional world. Concepts that are encountered in the real world, such as space, amount, size, location, and object, are transferred to user interface elements.

- *Structural metaphors* are characterized by the comparison of the structure of an object to another and are often rooted in the experience of the real world. For example, the concept of a real folder in a real office supports users in understanding that an icon of a folder as an interface element entails 'putting things into it' (in [BBN02], also referred to as *metaphorical entailment*).
- *Metonymy* is encountered when one element refers directly to its metaphorical function within the *metaphoric world* defined by its implemented metaphors. For example, using a magnification glass icon to describe a magnification interaction.

Marcus proposes a fifth type, *pragmatic metaphors*, that supports users in understanding the system's functionality by representing the abstract processes executed on the level of the operating system in a simpler form, effectively hiding complex processes behind simple interactions [Mar98]. The use of metaphors can lead to *metaphor mismatches*, in which the user falsely perceives a specific aspect of a system as related to some aspect from a different domain [CMK88]. However, the discovery of salient dissimilarities between domains can, just like salient similarities, help the user understand the system [CMK88]. Often, multiple metaphors are combined as *composite metaphor* to enhance functionality and divide tasks of a system into smaller units, for example, the desktop metaphor for personal computers can contain a calculator metaphor [CMK88]. In many cases, selecting appropriate metaphors is a key element to allowing novice users to perform tasks intuitively or, in some cases, at all.

Metaphors can be considered *novel metaphors* or *conventional metaphors* [BBN02]. *Conventional metaphors* are well-established metaphors that can be utilized by the target user group without much cognitive work, for example, WIMP systems in desktop computers. *Novel metaphors* are metaphors that are not *conventional*, characterized by not having been established in many systems and their unfamiliarity to users. Well-designed novel metaphors can become conventional if they are absorbed in the way developers and users think of systems. As new interfaces are often related to or even based on some pre-existing system or software, the metaphors of a predecessor system can also be utilized as conventional metaphors for the new system. [CMK88]. The best practices and related approaches to interaction, depending, for example, on the model of the user, goals in interaction, and new technological capabilities, form a "set of practices upon which a community has agreed," [SPR19] which is called a paradigm. New paradigms have been introduced that account for novel technological capabilities and new HCI developments, and effective interaction paradigms that make a system usable remain in the catalog of HCI. Today, there is not a single way of interacting with computer systems, but a collection of coexisting paradigms [WW11] that are further developed in research and development.

2.1.4 Usability and User Experience

One goal of interface research is to improve the usability of computer systems. A key concept in understanding usability is that users understand the interaction with their environment by constructing a *mental model* that represents the functionality of a system [CM88; Nor13; Ber+08] by "representing properties of the task environment which can serve the planning of activities and the control of acts when instantiated and activated by observation of the actual state of affairs" [Ras87]. Carrol defines the mental model to be "a rich and elaborate structure, reflecting the user's understanding of what the system contains, how it works, and why it works that way" [CO88]. It consists of a set of *concepts*, that describe categories and events with similar properties, *propositions*, that associate concepts with another, generalized knowledge in the form of *schemas*, and *scripts*, which describe events in a generalized way [Ber+08]. The representation of knowledge about interaction in the mental model allows for *skill-based behavior* (automated and subconscious appli-

cation of sensorimotor performances), *rule-based behavior* (the deliberate and consciously controlled selection of action based on stored procedural knowledge), and *knowledge-based behavior* (functional reasoning based on explicitly stored information about the system) [Ras87]. To facilitate the construction of a mental model, interfaces incorporate various design principles based on human psychology. Norman describes the "seven fundamental principles of design" [Nor13, p. 72] as follows:

- *Discoverability*: Discoverability refers to the degree to which users can easily locate and comprehend the available functions and features of a system or interface. It involves designing intuitive and transparent cues that enable users to quickly identify and understand the options and actions within a given context. Employing clear labeling, visual cues, and logical grouping enhances discoverability, promotes user engagement, and reduces cognitive load.
- *Feedback*: Feedback is the information provided to users in response to their actions or inputs. It serves as a mechanism for users to understand the outcome of their interactions and make necessary adjustments for future actions. Feedback can be described as the fundamental structure in all dynamic and cybernetic systems [For+68].
- *Affordance*: Weiser and Brown describe affordance as "a relationship between an object in the world and the intentions, perceptions, and capabilities of a person" [WB96]. The term affordance has its roots in ecological psychology. It was introduced as a term to describe perceptual cues in an environment that guide the actions of an agent [Gib14]. The term was later popularized in the context of computer interfaces to describe the properties of objects and provide users with information on how to use them [LJ+17].
- *Signifiers*: Signifiers are closely related to affordances. Signifiers communicate to a user the "purpose, structure, operation, and behavior of an object" [Jer15, p. 279] and make an intended affordance visible. An example is an interactive touch screen with a button. Whereas the entire screen affords the action of touch, the button signifies where the touch event invokes an intended action. Signs and perceptual perceptions, such as visual or auditory cues, provide indications about the available actions or functions. They are intentionally designed to help users understand and purposefully utilize the affordances of an object or environment.
- *Mapping*: Mapping refers to the relationship between controls or actions and their effects. A well-designed mapping that implements a "compliant" [Jer15, p. 282] spatial layout and temporal consistency ensure that users can predict the outcomes of their actions.
- *Constraints*: Perceivable constraints can facilitate the understanding and interpretation of an interface by limiting the possible interactions. Constraints can be physical, logical, semantic, and cultural.
- *Conceptual model*: The conceptual model describes the functions of a system from the designer's perspective. By relying on the presented principles of interface design, the conceptual model can be aligned with the user's mental model. If the working mechanisms of interactive systems are not clearly conveyed, users can be forced to create individual models based on observations or previously encountered similar systems. These user-created models can differ from the *intended user's model* [CM88], which may lead to false actions [Ber+08].

One important factor in creating usable systems is learnability [NM90; ISO22]. However, the term 'learnability' is not precisely defined and often used without further explanation [GFA09], but it can generally be paraphrased as "acquiring knowledge and skills and having them readily available from memory so you can make sense of future problems and opportunities" [BRIM14], often specifically in the case of novice users encountering an unfamiliar system [GFA09]. In a further definition by Nielsen, 'learnability' specifically means the "novice's user experience" when first interacting with an unknown system [Nie94]. In ISO/IEC 25010, learnability is defined as the "[d]egree to which a product or system can be used by specified users to achieve specified goals of learning to use the product or system with effectiveness, efficiency, freedom from risk and satisfaction in a specified context of use" [ISO22]. Preece et al. distinguish between five heuristic strategies that are applied when previously unknown interfaces are encountered [PBU93]:

- *Learning through doing.* Users try out different ways of interaction without any further knowledge, possibly skipping instruction manuals or other time-consuming resources (with a high chance of failing).
- *Learning by active thinking.* Users try to predict system behavior and derive the next step in a reasoning process from observations.
- *Learning through goal and plan knowledge.* Users infer the next steps in interaction by performing a means-end analysis and optimizing the discrepancy between the target and current state, moving forward from sub-goal to sub-goal.
- *Learning through analogy.* Users transfer knowledge from a well-known interaction to a novel interaction.
- *Learning from errors.* Negative feedback from making mistakes can guide users similarly to positive feedback if it is conveyed appropriately.

Different interface implementations can be objectively evaluated regarding their usability. Nielsen and Molich define heuristics to evaluate the use of dialogue-based systems, which are based on human capabilities and behavior, such as a consistent system design, providing appropriate feedback, and requiring only low cognitive resources [MN90; NM90]. Quesenbery defined the *5 E's* of usability engineering [Que04]: *effective*, *efficient*, *engaging*, *error tolerant*, and *easy to learn*. Other usability goals have been defined as *memorability, safety*, and *utility* [PSR15]. Beyond the Taylorist perspective that aims primarily to increase the efficiency in performing a task [Dix+04], other important qualities of interfaces and products have been incorporated into research, such as aesthetics, emotion, usefulness, meaning, entertainment, and satisfaction [PSR15; HP12]. Interfaces not only possess a *pragmatic* quality that aims at allowing users to efficiently achieve goals but also a *hedonic* quality that describes the feeling and experience of using a specific product [HBK03]. These experiential factors are often referred to as user experience (UX). Hassenzahl et al. identified eight psychological needs that can be addressed in products that are based on technology to provide a positive UX: *security*, the striving for *meaning*, the feeling of *competence* and *autonomy*, *relatedness* to others and *popularity*, *stimulation*, and *physicalness* [HDG10].

2.1.5 Natural User Interfaces

A key idea in usability engineering and UX design is aiming at human capabilities to make the interaction less artificial and computer-oriented, which enables a more 'natural' interaction with computer systems. Although the term 'natural' has often been used in HCI research to describe interfaces that are tailored for the human user, the precise meaning is still debated. The use of this term has even been criticized as a "hype term" [Ort+16, p. xxxiii] or "good for marketing" [O'h+13]. In 2010, Don Norman published an article

titled "Natural User Interfaces Are Not Natural" [Nor10], in which he especially criticizes the marketing of gestural interfaces as a universal means of natural interaction.

Various authors have proposed descriptions or definitions of the constituents that make a user interface 'natural'. One of the first instances of explicitly using the term "natural user interface" is a PHD thesis from 1973 by De Fanti on providing, what he called, a *habitable* (in reference to Watt's *habitability*[2] of human-computer languages [Wat68]) system for computer animation [DF73]. In his thesis, the key characteristic of a habitable system and, subsequently, of a natural user interface, was creating a system "whose conventions are easy to learn and whose commands are powerful, yet easy to use" [DF73, p. 24]. Wigdor and Wixon describe the term "*natural* referring to the way users interact with and feel about the product, or more precisely, what they do and how they feel while they are using it" [WW11, p. 9]. Hansson et al. consider the interaction with computer systems 'natural' if it intuitively appeals to the user, facilitating the transfer of knowledge and skills acquired from familiar environments and past experiences [HWS97]. Similarly, Blake defines NUIs as: "A natural user interface is a user interface designed to reuse existing skills for interacting directly with content" [Bla11, p. 2]. In his definition, he focuses on the three aspects [Bla11]: 'natural' interfaces are i) *designed*, which emphasizes the importance of intentional considerations regarding the user group, content, and use context. ii) They *reuse skills* that are acquired in real-world human-environment interaction. iii) *Direct interaction with content* is the primary interaction mode in 'natural' interfaces. The reuse of skills as an aspect of natural interaction can alternatively also be called 'intuitive' as defined by Mohs et al.: "A technical system is intuitively usable if it leads to effective interaction through the unconscious application of prior knowledge by the user" [Moh+06]. Overall, 'natural' can, in today's use, be seen as an umbrella term referring to various aspects of interfaces, including nature-inspired design, systems that conform to expected behavior, and intuitive use [Hir+22].

2.2 Virtual Reality

VR systems have received major interest in recent years. The increased technological capabilities allow for a wide distribution of VR devices outside of industry and research facilities, making VR an interesting contemporary approach in HCI with a wide range of potential applications. This section provides an overview of important topics in VR research from the conceptual and technical perspectives, with a focus on interaction in virtual environments.

2.2.1 Reality-Virtuality Continuum

The term 'virtual reality,' as it is used today, was coined by Jason Lanier in the 80s, who described VR to be "a technology that uses computerized clothing to synthesize shared reality. It recreates our relationship with the physical world in a new plane, no more, no less. It doesn't affect the subjective world; it doesn't have anything to do directly with what's going on inside your brain. It only has to do with what your sense organs perceive" [Lan88]. Various scientific definitions exist that describe what 'virtual reality' exactly encompasses. Jerald, for example, defines virtual reality generally "to be a computer-generated digital environment that can be experienced and interacted with as if that environment were real" [Jer15]. Brooks lists four main and four auxiliary technologies that are required for such a system: The main technologies are (1) visual and aural displays, (2) a computer graphics rendering system, (3) a tracking system for head and limbs, and (4) a database

[2] Watt defines a habitable language in a dialogue-based computer system as such when "it's users can express themselves without straying over the language's boundaries" [Wat68]

system that provides models of the virtual environment. The auxiliary technologies are (5) spatial sound, (6) simulation of haptic forces, (7) systems for user input, and (8) interaction techniques [Bro99]. Besides 'virtual reality,' the synonymous term [Luc07; LJ+17] (immersive) virtual environment is used to describe a computer-generated world that is "capable of delivering an inclusive, extensive, surrounding and vivid illusion of reality to the senses of a" a user [SW97]. In this definition by Slater et al., *inclusive* is defined as allowing a user to shut out reality while engaging with the system, *extensive* describes the range of sensory modalities (vision, auditory, tactile, etc.), and *vivid* describes the fidelity of artificial stimuli regarding reality.

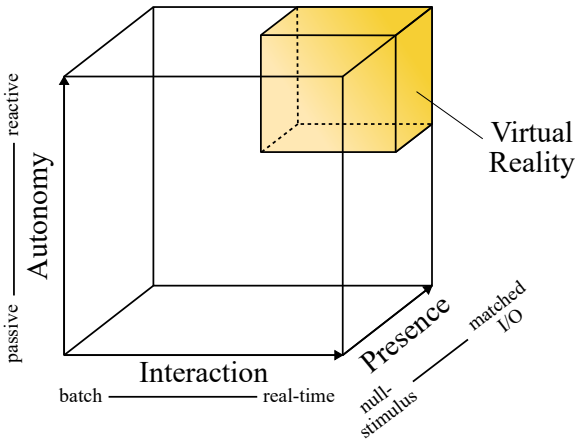


Figure 2.4: The three-dimensional AIP-cube describing virtual reality proposed by Zeltzer. (Figure based on [Zel92]).

Zeltzer describes the AIP-cube to distinguish between telepresence systems and screen-based graphical applications [Zel92] using the orthogonal axes *Autonomy*, *Interaction*, and *Presence*. Autonomy describes the ability of the environment and objects to react to user input and events. A static environment that does not react to the user, e.g., static 3D objects, exhibits no autonomy, whereas objects and agents that show knowledge-based behavior increase autonomy. Interaction is a dimension of the degree of control over variables by the user, with "no interaction" on one side and fully comprehensive access to all parameters on the other. Presence describes the sense of 'being there' due to perceiving high-fidelity sensory input and output from the system that matches the sensorimotor skills of users. When a system is not capable of generating a system response to sensorimotor skills performed by the user, for example, shifting the viewport according to head movements, this would be considered a *null-stimulus* [Zel92], and the achievable presence would be zero. A full match of system capabilities and all input and output of human users corresponds to a high presence within the environment. A system with high autonomy, high interaction, and high presence would be a perfect virtual reality system, the "grail" [Zel92] for VR research.

Other terms exist that describe technology related to VR, for example, mixed reality (MR). In 1988, Lanier described MR in relation to VR as a "more sophisticated [virtual environment] where you can blend virtual objects and physical objects so that you can live in a mixed reality" [Lan88]. A common definition in research that further expands this description is the Reality-Virtuality Continuum by Milgram and Kishino [MK94], published in 1994. In this continuum, one extreme is the real world without virtual content (*Reality*), and the other is a perfect "matrix[3]-like" [SSW21] virtual environment (*Virtuality*) that is equivalent to Sutherland's *Ultimate Display* [Sut65] in which users would not be able to

[3] Referencing the 1999 movie The Matrix.

distinguish between reality and virtual content. Between the extremes *reality* and *virtuality* lies mixed reality, which is further divided into *augmented reality* and *augmented virtuality*. *Augmented reality* describes extending our perception of reality by including computer-generated virtual components into reality. Three key requirements for AR systems have been proposed by Azuma [Azu97]: They need to be (1) *interactive in real-time* and (2) *combine virtual and real content*, which is (3) *registered in 3D* in the real world. *Augmented virtuality* refers to the augmentation of virtual environments with real-world content. In analogy to Azuma's definition of AR [Azu97], systems can be considered AV if they combine 3D-registered virtual and real content in a real-time interactive IVE. Another term is *Extended Reality* (also *Cross Reality*) (XR) that describes, in a common interpretation [Doe+22; Rau+22; Bal+21], the union of virtuality, VR- and MR-systems and its subclasses AR and AV, making the 'X' a placeholder. The relationship between the terms, as understood in this thesis, is depicted in Figure 2.5.

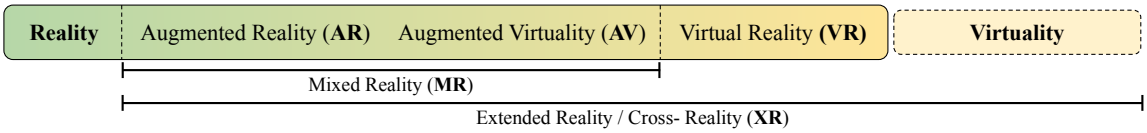


Figure 2.5: Relation of common terms in mixed-reality research based on the Reality-Virtuality Continuum [MK94].

2.2.2 Immersion, Illusions, and Qualia

A key element that distinguishes VR from other types of media, such as books, theater plays, or movies [Sch08] is immersion. In the context of VR research, immersion describes the system's ability to provide *sensorimotor contingencies* to a user [Sla09]. These include *valid actions*, which allow users to meaningfully change their perception, for example, moving their head to shift the viewport (*sensorimotor actions*), and also actions that affect the virtual environment (*effectual actions*) [Sla09]. For simplicity, in this thesis, *immersion* is treated as synonymous with the technical *system immersion*, following the definition by Nilsson et al. [NNS16], and does not include *challenge-based immersion* and *narrative immersion*. Immersion can lead to a feeling of being transported to a different location, the "sense of being there" [IJs05]. Marvin Minsky introduced the term tele-presence in 1980 in the context of operating robots from afar [Sla09] and emphasized the challenge of coupling artificial devices in a natural and comfortable way with the human sensory system to achieve this experiential effect [Min80]. In today's VR research, this is typically referred to as the feeling of *presence*.

According to Skarbez et al., presence is a construct emerging from different experienced illusions: *Place illusion*, *copresence illusion*, and *plausibility illusion* [SBW17] (see Fig. 2.6). Presence is related to immersion in the sense of immersion "*provid[ing] the boundaries within which [Place Illusion] can occur*" [Sla09]. Slater defines place illusion as "*the strong illusion of being in a place in spite of the sure knowledge that you are not there*" [Sla09]. Plausibility illusion describes, according to Slater, "the illusion that the scenario being depicted is actually occurring" [Sla09]. This is, according to Skarbez et al., depending on the objective property *coherence* [SBW17], which is defined "as the set of reasonable circumstances that can be demonstrated by the scenario without introducing unreasonable circumstances, where a reasonable circumstance is a state of affairs in a virtual scenario that is self-evident given prior knowledge" [SBW17]. For VR environments that include other users or simulated intelligent human characters, *social presence* is a further influential factor that is dependent on the design and behavior of the depicted avatars.

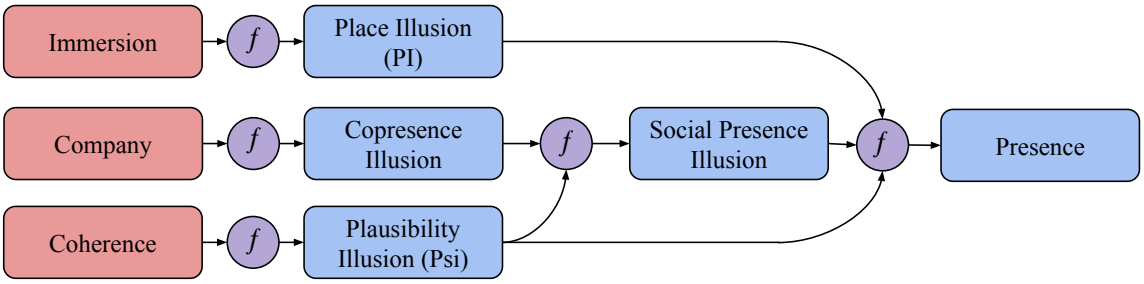


Figure 2.6: Reproduction of the *Presence Model* presented by Skarbez et al. [SBW17] with objective factors (red), experienced qualia (blue), and functions based on individual differences (violet).

The feeling of presence is a central quale in VR research. In general, a quale describes a subjective experience, such as perceiving the color red or feeling happy. Qualia are ineffable, intrinsic, private, and direct experiences [Den88]. In VR research, qualia refer to subjective experiential effects, which are introduced by the content and properties of an IVE. Often, users are, on a cognitive level, well aware that a VR experience is not real but still respond to it on a behavioral and physiological level, similar to a real-world experience, for example, in virtual height experiments [Uso+99; Mee+03]. These illusions "occur when media instrumentation stimulates neural bottom-up multisensory processing, sensorimotor self-awareness frameworks, and cognitive top-down prediction manipulations and furthermore allows these to reconcile in such a way that semantic violations are infrequent" [GFL17]. In contrast to technical properties, such as immersion, or quantitative measurements, such as task performance time, qualia cannot be observed directly. Instead, they have to be measured using adequate tools such as questionnaires. In the case of *presence*, for example, several questionnaires exist that are intended to capture the experience of presence [Sch+19b], e.g., the Spatial Presence Experience Scale (see appendix F).

Another important quale is the *sense of embodiment*. Kilteni et al. define "[the sense of embodiment] toward a body B is the sense that emerges when B's properties are processed as if they were the properties of one's own biological body" [KGS12]. Furthermore, they analyze three underlying sub-components: the *sense of self-location* (the spatial experience of being inside a body), *sense of agency* (global motor control including the subjective and conscious experience of action and will), and *sense of body ownership* (self-attribution of the virtual body as the source of sensations). The experience of embodiment is maximized when all subcomponents are maximized [KGS12].

2.2.3 Realism and Interaction Fidelity

Sutherland imagined in *The Ultimate Display* the possibility of creating computer-generated environments in alignment with how humans perceive the physical reality [Sut65]. Such a perfect copy of reality can be analyzed on two levels: At the lower sensation-perception level, all sensorimotor skills [Zel92] and sensory modalities of the human user have to be addressed *extensively* with a *vivid* [SW97] quality, that cannot be distinguished from real-world sensations. At the higher cognitive level, the properties of the conscious experience can be aligned with experiencing the real world. Realism is, however, hard to capture fully. In the context of digital games, for example, realism can be considered a complex multidimensional construct that encompasses, among other factors, perceptual aspects as well as historical realism and the realism of avatars [Rog+22]. In VR research, diverse terms are used to describe realism, for example, *congruence* ("To what extent was what you experienced in the virtual world congruent to other experiences in the real world?") [Bañ+00], *verity* ("the degree that our natural environment is represented in a virtual environment")

[TM94], or *external plausibility* ("how consistent the virtual environment is with the users' real-world knowledge") [Hof+20]. Chalmers distinguished between *Is/Not Believable* on the one hand and *Is/Not Physics* on the other which both need to be sufficiently targeted to reach a *there-reality* in which "the same perceptual response from a viewer as if they were actually present, or "there", in the real scene being depicted [is evoked]" [CF08].

Realism can be considered a form of quale, as it involves subjective experiences related to the perception of reality. Several questionnaires have been developed to quantify how individuals perceive and experience realistic environments or situations in IVEs. Examples are the German VR realism scale [PD13], which focuses on visual realism and human avatars, the IPgroup questionnaire (IPQ) [SFR01], which contains questions targeting the subjective impression of the overall realism of a virtual scene in comparison to the real world, and the feeling of being present and involvement in the virtual scene. In some cases, the feeling of presence and perceived realism are treated similarly, as both concepts relate to an individual's immersion and engagement with a given environment. In early experiments, the realistic behavior of objects and natural movements of the virtual body were considered important factors influencing the feeling of presence [SU93]. The Spatial Presence Experience Scale (SPES) [Har+15], for example, contains the dimensions *self-location* and *possible actions*, which implicitly rely on the replication of real-world actions in the virtual environment. However, an important finding for VR is that a high feeling of presence can be achieved regardless of the *external plausibility* [Hof+20] as long as the system provides a *coherent* [SBW17] experience with a high *internal plausibility* [Hof+20]. Plausible scenarios have an impact on affective responses to virtual experiences [New+22; Gis+19].

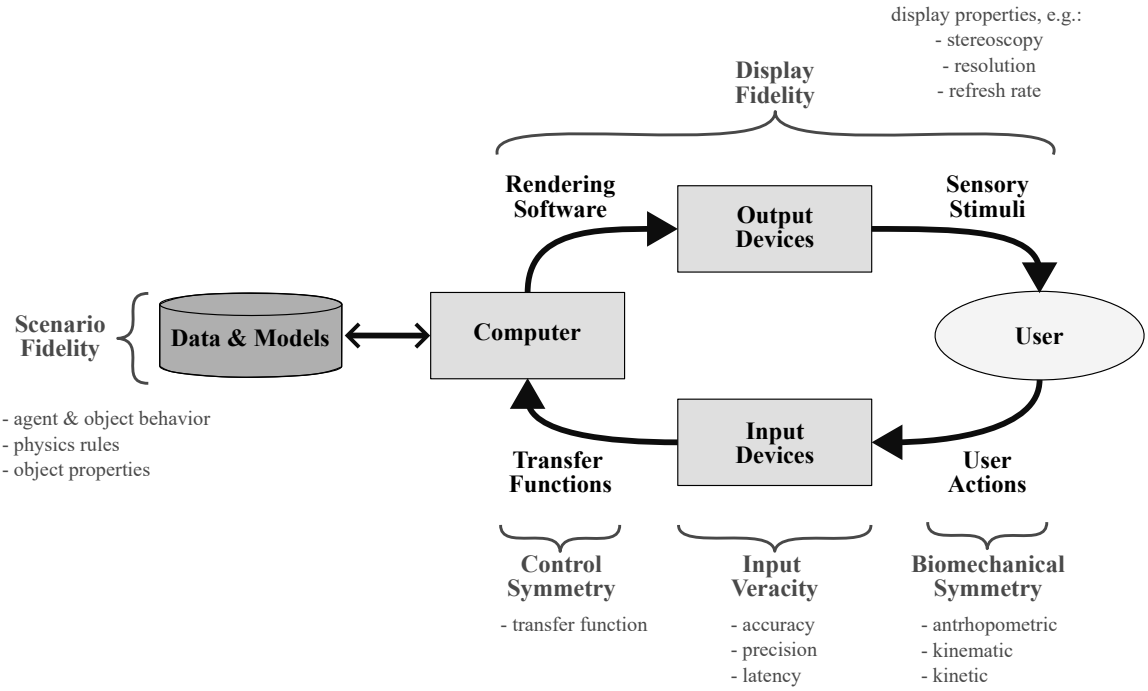


Figure 2.7: McMahan's User-System Loop describing the symmetry of interaction between a user and the computer system via input and output devices. Based on [MLP16; Rag+15].

An important model in VR research that focuses on the realism in the interaction loop between a user and a VR system has been proposed by McMahan et al. [MLP16]. The Framework for Interaction Fidelity Analysis (FIFA) (see Fig. 2.7) follows in its structure the interaction framework proposed by Abowd and Bale [AB91] (see Fig. 2.1). It can be utilized to objectively analyze the degree of exactness of recreating the real-world interac-

tion, also referred to as *fidelity*. McMahan describes three main components for interaction fidelity [LJ+17; McM11]: First, *biomechanical symmetry* which describes the exactness of reproducing the real-world body and movements with the sub-components *anthropocentric symmetry* (reproducing the human body), *kinematic symmetry* (reproducing body movements) and *kinetic symmetry* (reproducing forces). Second, *input veracity*, as the technical capability of a system to capture user input with the sub-components *accuracy*, *precision*, and *latency*. The last main component is *control symmetry* with one sub-component that describes the transfer function of the interaction technique to the affected physical properties of the real-world counterpart action (*transfer function symmetry*) [LJ+17]. Besides interaction fidelity, Ragan et al. expand FIFA with display fidelity and scenario fidelity as components of analysis [Rag+15]. Visual display fidelity includes vision-related aspects, such as field of view, display resolution, and frame rate, whereas scenario fidelity encompasses the behaviors of agents, properties of objects, and implementation of real-world physics rules [Rag+15; LJ+17].

However, providing a realistic environment is not the only way of implementing interaction in IVEs. XR technology also enables behavior and actions that are not based on reality. One motivation for creating non-realistic environments is presented in Jacob's general framework for reality-based interaction [Jac+08a]. In this framework, it is proposed that certain themes of the interaction between humans and the real (non-digital) world can be addressed in post-WIMP interfaces. First, *naïve physics*, the fundamental understanding of basic physical laws, such as gravity. Second, *body awareness & skills*, the perception of one's physical body and how basic sensorimotor actions can be used to produce movements. Third, *environmental awareness & skills*, in which users perceive their environment and know how to interact with it. And last, *social awareness & skills*, the awareness of other people in their environment and knowledge about interacting with them. These four themes can be present in an artificially computer-generated environment, or they can be reduced to enhance specific properties of the interface. Jacob describes six different categories: *expressive power* (the extent of possible actions), *efficiency* (speed of performing tasks), *plasticity/versatility* (quantity of tasks from different domains), *ergonomics* (interaction in accordance with the human body), *accessibility* (including users with different abilities), and *practicality* (considerations for the development and production). The key idea in Jacob's framework is that the design of systems can include tradeoffs between realism (in regard to the real non-digital world) and enhanced functionality (depicted in Fig. 2.8). In the context of VR interaction, the idea of reducing realism to enhance functionality has specifically been expressed by several authors, for example, Kulik [Kul09] and Bowman et al. [BMR12].

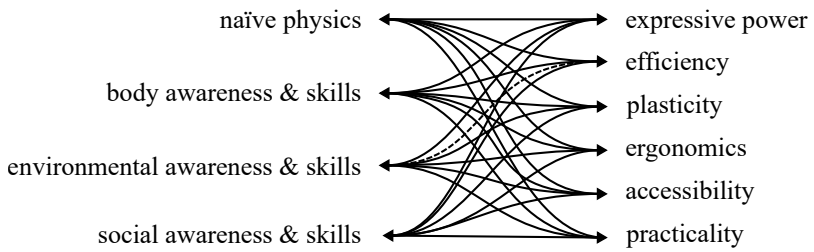
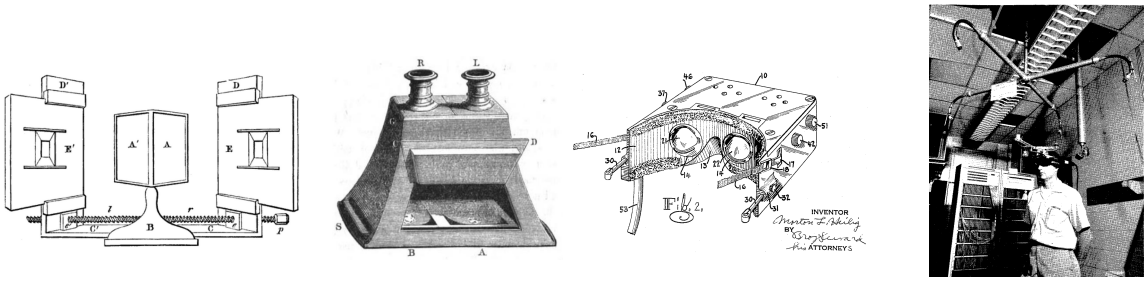


Figure 2.8: The tradeoffs between realism and desired properties proposed by Jacob (Figure taken directly from [Jac+08a]).

2.2.4 VR Systems

2.2.4.1 Historic Foundations

From a technical perspective, the foundations of today's VR technology and stereoscopic displays were laid out in the 19th century. The Wheatstone Stereoscope [Whe38] from 1838 and the Brewster Stereoscope from 1849 [Far94] were experimental systems for researching stereoscopic vision. Both systems provide individual visual perceptions to each eye that produce a 3D effect when perceived. In 1896, Stratton invented upside-down goggles, which inverted the viewing direction, to investigate stereoscopic vision further. He reported in a self-experiment that objects were perceived "to be real things, like the things we see in normal vision, but they seemed to be misplaced, false, or illusory images between the observer and the object" [Str96]. In 1960, Morton Heilig received a patent for a *stereoscopic-television apparatus for individual use*, also called *Telesphere Mask*, which enabled the display of stereoscopic video content for each eye, stereo audio, and even the production of synchronous olfactory stimuli. Although the technical drawings of the *Telesphere Mask* resemble today's VR devices, Heilig's device did not provide sensorimotor contingencies such as head tracking, nor interactivity, nor was the content computer-generated.



a) b) c) d)

Figure 2.9: Historic evolution of stereoscopic displays. From left to right: Wheatstone Stereoscope (1838) (a), Brewster Stereoscope (1849) (b), Heilig's Telesphere Mask (1960) (c), and Sutherland's Head-Mounted Display (1968) (d).

Figures adapted from [Bre56] (a, b), [Hei60] (c), [Sut68] (d).

The first actual system that was capable of displaying computer-generated content that could be viewed using head movements was the *Head-Mounted Three-Dimensional Display*, a device developed by Sutherland in 1968 [Sut68]. This system, often referred to as *Sword of Damocles*[4], was able to display wireframes that were superimposed onto the view of the real room using half-silvered mirrors in a head-mounted display (HMD). Regarding this system, Sutherland said in 2005 that the computer systems used to produce the visual output "were not up to the task of rotating the image, let alone rendering it in any semi-realistic form" [Com14]. These early and subsequent VR systems were expensive, and the rendering and interaction capabilities were limited compared to today's standards, which prevented the widespread use of this technology [Pau91; Wal90]. Still, in the early 1990s, researchers were conscious of the limited technological capabilities of VR systems and the inherent impact on applications [Zel92; VHS94; You+96]. Youngblut et al., for example, stated in 1996 that "in no instance ... the interface technology match[es] human capabilities for the relevant sensory modality" [You+96]. In 1994, Veron et al. formulated *realistic goals* for usable HMDs: a refresh rate of 70 Hz, a resolution of 2 arc minutes per pixel (corresponding to 30 pixels per degree, a field of view of 140°, a binocular overlap of 70°, a 15 bit RGB color range, no tether, and a weight below 340 g [VHS94]. Beginning with the Oculus

[4] This name was initially only describing the mechanical tracking system mounted to the ceiling, however, in literature, the term is often used to refer to the system as a whole [Doe+22].

Kickstarter campaign in 2012 [Kic16], the new hardware initially developed for mobile devices (e.g., high-resolution displays, spatial tracking, battery capacity) [Ste16a] enabled the production of much smaller VR devices at lower prices with better resolution. These modern HMDs almost meet, or, in some cases, even exceed the requirements proposed by Veron et al. [VHS94] (see Table 2.1) and, finally, enable the use of VR and AR technology in consumer and industry scenarios at affordable prices.

2.2.4.2 Display Devices

Today, a wide range of immersive systems exists that are capable of displaying VR or AR content. Two important system types that are capable of displaying virtual content can be distinguished: stationary systems and HMDs [Doe+22]. Stationary systems use physical objects in the environment of the user to display virtual content. CAVE systems, as an example of stationary systems, project digital content that corresponds to the user's perspective onto the walls, ceiling, and floor of a room to provide real-time computer-generated visual content to a large portion of the visual field [CN+92]. Similarly, immersive domes display content on a user-surrounding hemisphere [REN89]. Fishtank VR systems [WAB93] and workbenches [LJ+17], on the other hand, only change a small area of an object to provide a 'window' to the virtual environment. Fishtank VR and workbench systems are, therefore, often classified as *semi-immersive* [BC03] as their provided sensorimotor contingencies are limited to a specific area of visual perception. HMDs, which are directly located in front of the user's eyes, and CAVE systems, which surround the user, on the other hand, exclude external perception of the real world, which makes them fully *immersive*. The focus of this research work is HMD-based VR, so only HMDs are further described in the remainder of this section.

Figure 2.10: The tethered system HTC Vive (2016) with lighthouse 1.0 base stations and controllers.

In order to allow users to form a three-dimensional experience of a computer-generated environment that corresponds to the spatial experience of the real world, the HMD's display provides diverse visual cues (*stereopsis*, *monocular*, *oculomotor*, and *motion parallax* [LJ+17]). For *stereopsis*, a stereoscopic view of a virtual scene with a distinct virtual camera position for each eye is rendered. The binocular disparity, the difference between both perceptions, is processed in the brain to estimate 3D depth in analogy to how 3D perception is achieved in the real world. Monocular clues are, for example, occlusion, perspective, and the relative size of known objects. Oculomotor cues are *accommodation*, the adjustment of the lenses in the eye to focus on a specific distance, and *vergence*, the rotation of the eyeballs to focus on a point in the distance. Additionally, *motion parallax*, which is

perceivable as the difference in speed of objects moving in relation to the distance from the viewer, provides cues for 3D information to a user in motion (such as self-motion or vehicular motion, for example, in a car). The displays used in HMDs have different characteristics that further influence the quality of visual perception. Important factors besides the ability to provide depth cues are the field of view, spatial resolution, screen geometry, light transfer, refresh rate, and ergonomics [LJ+17]. Current-generation consumer-level devices offer a FOV of approximately 100° horizontally and vertically, 20 pixels per degree, and a refresh rate between 90 and 120 Hz (see Table 2.1), which is considerably lower than human vision, but sufficient for creating immersive experiences.

Depending on the requirements of a VR application, the rendering of virtual content can either be performed on a powerful workstation in a tethered setup or, if the requirements for computation and rendering are lower, on the HDM itself as a standalone device. In the case of a tethered setup, the HMD is connected to the workstation via a long, multicore cable that transmits the rendered virtual content as a video signal, provides electrical power to the HMD, and receives tracking data from sensors located within the HMD. Standalone devices (for example, the Meta Quest 2, see Fig. 2.11a) are battery-powered and perform all calculations on the device, which limits the capability for simulating and rendering content. However, the enhanced freedom of movement without consideration for a cable in the physical non-virtual environment and the enhanced mobility can, in some cases, be preferable over rendering power. A subclass of HMDs are see-through AR devices (for example, the Microsoft HoloLens 2, see Fig. 2.11b), which superimpose virtual content onto the perceived real world. See-through AR HMDs can be video-based, using low-latency cameras and displays, or optical see-through using optical elements such as prisms, semi-transparent mirrors, and waveguides [Xia+19] to display virtual content [Doe+22].

(a) Meta Quest 2 (2020). (b) Microsoft HoloLens 2 (2019).

Figure 2.11: Two examples of state-of-the-art standalone HMDs.

Besides visual content, other human senses can also be targeted with output devices. Often, HMDs include stereo loudspeakers or plugs for headphones to provide 3D acoustic stimuli. Haptic displays have been developed for industry and research purposes [LJ+17] that provide the perception of force or resistance when a virtual object is touched. Some of these systems are *ground-referenced* and stationary, others are body-referenced and placed on a specific body part. Some sophisticated data gloves can simulate haptic resistance for each finger individually [Doe+22]. However, these haptic devices are expensive and cumbersome, making them more of a niche product for sophisticated applications. Consumer-oriented HMDs, on the other hand, are often delivered with a pair of system-specific generic input devices that provide a combination of buttons, triggers, and a touchpad or joystick (see, for example, Fig. 2.10). Some of these input devices can provide a high-frequency (around 1000 Hz) vibration output to simulate an artificial vibrotactile sensation. The neuroplasticity of the brain allows humans to adapt to this change and perceive vibrations

as a substitute tactile stimuli for realistic haptic sensation after some time of adapting [LJ+17]. Other modalities, for example, olfactory and gustatory displays, or simulation devices for temperature, are less common and can mostly be considered research-oriented prototype systems [LJ+17]. In this research work, only visual and auditory content are investigated.

2.2.4.3 Tracking

Head Tracking Head tracking is a fundamental component of VR systems. Head movement-induced shifts of the virtual viewport correspond to the egocentric experience of moving one's head in the real world, which forms a fundamental sensorimotor skill in exploring a physical three-dimensional environment. Head tracking requires a tracking system with high frequency (> 90 fps, the same as the HMDs display refresh rate), low latency (< 20 ms), high accuracy (< 1 mm), low noise, and low drift [Doe+22; Cue17]. Often, a tradeoff between cost and quality has to be considered, but other factors, such as the number of unique tracked devices, sensitivity to external influences, calibration processes, and usability, have to be considered as well [Doe+22]. For contemporary commercial systems, an important way of tracking head and controller movements is a hybrid tracking using high-frequency internal measurement units (IMU) within the HMD to track head rotations and acceleration, combined with low-frequency optical pose tracking.

A common approach for optical tracking is based on infrared (IR) light emission and detection. The Vive system by HTC, for example, uses external IR light-emitting tracking stations, so-called *lighthouses* (see Fig. 2.10), and photoelectric sensors built into the HMD [NCL17; HAC16] to track the HMD. In this setup, the lighthouses first send laser pulses for synchronization, followed by a directional laser sweep. The first pulse starts a timer for each sensor in the HMD, and the time difference between the laser pulse and the laser sweep enables the calculation of the distance from each sensor (time-of-flight) to the tracking station. The HMD's pose can then be estimated by combining the information from all sensors. For synchronization, the tracking station, the HMD, the controllers, and other devices communicate via Bluetooth. The Oculus CV1 utilizes a different approach for IR-based tracking, in which the HMD is equipped with a unique pattern of precisely spatially arranged IR-emitting LEDs that is detected by external tracking stations in the so-called *constellation tracking system* [Fel15] to estimate the pose of the HMD. Although IR-based tracking has some benefits over other methods, such as being usually reliable, sufficiently accurate, easy to set up, and cheap, it requires an instrumented space and additional hardware. An alternative that gained popularity in consumer devices in recent years exploits inside-out camera tracking, which calculates the pose change of the HMD using a vision-based detection of features within the real environment and estimates positional changes from the optical flow. Dedicated cameras record the real environment to extract feature points, which are then continuously tracked, an approach called SLAM (simultaneous localization and tracking). This approach is used in many current-generation devices, such as Meta Quest 1, 2 & 3, Hololens 1 & 2 [Doe+22], PlayStation VR 2 [Pla23], and Apple Vision Pro [App23]. A benefit of these systems is a very simple setup, which requires no additional hardware and instrumentation.

Hand Tracking For interactive purposes, the body of the user can be tracked. The hands, as central body parts for interaction with the real world, have received major interest in research and development. Together with the Eyephone and the DataSuit in the initial catalog of products by VPL in the 1980s, so-called DataGloves provided the means for interacting within a virtual environment [Bla+90]. Data gloves, motion tracking, and vision-based systems are important contemporary approaches for fully articulated finger

tracking. All these systems exhibit varying advantages and disadvantages, thereby presenting distinct characteristics and drawbacks that must be considered in system design, for example, accuracy and precision, technological properties (drift, noise, latency), comfort, model precision, and distribution [DGS20].

Figure 2.12: A collection hand tracking systems [DGS20]. From left to right: Senseglove, AvatarVR, Sensoryx VRfree, IR marker tracking, Meta Quest 1, HoloLens 1, Leap Motion.

Data gloves can exploit various technologies, such as IMUs or electrical bend sensors. One of the first commercially available data gloves was presented by VPL in 1987 [Yan+19]. The advantages of glove-based systems are the direct measurement of angles with a high sampling rate, no occlusion, and relatively low-cost technology [HS14]. However, gloves are typically cumbersome and require instrumentation and calibration [HS14], which can negatively impact the user experience. *Motion tracking systems* employ an array of IR cameras and retroreflective markers or active LEDs to reconstruct the 3D position of joints in a defined tracking volume. Commercially available systems, such as Vicon and OptiTrack, enable the real-time calculation of 3D position for multiple *rigidbodies*, a set of rigidly connected spatially arranged markers, with a high frame rate and high precision, typically less than one millimeter. Systems can be extended with more cameras to reduce occlusions and enhance precision. A drawback of these systems is the high cost and limitation of tracking to the static tracking volume. *Vision-based systems* use generative (based on, e.g., a defined skeletal model and information about previous hand poses) or discriminative (machine-learning-based transformation of input data, e.g., a depth image, to joint positions) approaches to reconstruct hand and finger information from RGB or depth camera image data [Hua+21]. Some contemporary standalone HMDs, such as the Meta Quest 2, are shipped with camera-based hand tracking, which reconstructs finger flexions and hand positions from monochrome images using machine learning [Han+20]. The biggest drawback of vision-based systems is the loss of tracking due to occlusions. Missing parameters are often inferred or discarded, and interaction techniques are limited to what can reliably be tracked [HS14]. Although huge steps have been made in the past towards vision-based tracking systems using machine learning and computer vision [Olv+22], the performance of today's systems needs to be improved to allow a thoroughly reliable and articulated hand tracking [Hua+21], so, today, hand interaction using this approach is often limited to a small number of basic interactions, for example, tracking the three-dimensional position and orientation of the hand and detecting 'pinch'-gestures in which the thumb and index finger are tapped quickly.

2.2.4.4 Current-Generation Examples

Current-generation devices have reached a quality that allows for a wide range of applications at a low price. They are by no means perfect in creating the illusion of a full virtuality, but their capabilities are sufficient to produce visual and auditory stimuli that easily produce a feeling of presence. In this thesis, four different representative devices for the years 2017 - 2020 were used for research (two tethered HMDs and two standalone

HMDs). For reference, important specifications[5] [VRC23] are presented in the following table 2.1.

	Odyssey	Vive Pro Eye	Quest 2	HoloLens 2
Manufacturer	Samsung	HTC	Meta	Microsoft
Type	Tethered VR HMD	Tethered VR HMD	Standalone VR HMD	Opt. see-through AR HMD
Year of release	2017	2019	2020	2019
Resolution per eye	1440 x 1600 px	1440 x 1600 px	1832 x 1920 px	1440 x 936 px
FOV (h, v)	101°, 105°	98°, 98°	97°, 93°	43°, 29°
Optics	Fresnel	Fresnel	Fresnel	Waveguide
Refresh rate	90 hz	90 hz	120 hz	90 hz
Tracking	2x inside-out camera-based	2x HTC Vive base stations 2.0	4x inside-out camera-based	4x inside-out camera-based
Input modalities (built-in)	Controller, voice	Controller, voice eye-tracking	Controller, voice hand-tracking	Hand-tracking, eye-tracking, voice

Table 2.1: Specification of VR and AR HMDs that were utilized in this thesis.

2.2.5 3D User Interfaces

Traditional user interfaces can be considered one-dimensional (CLIs) or two-dimensional (GUIs) [Nie94]. Instead of utilizing the human's physical body or fully employing the natural abilities of interaction with the real world, artificial ways of interaction have been created to interact with such systems. IVEs, on the other hand, enable full three-dimensional interaction that resembles interaction in the real world, and so-called 3D user interfaces (3DUI)s are a dominant way of interacting in VR [LJ+17]. In these interfaces, interaction takes place in a 3D space, and it exploits the human body and spatial capabilities.

However, 3DUIs are not limited to replicating real-world interaction in the virtual environment. Transferring real-world interaction to an IVE and, thereby, limiting interaction to familiar ways of performing tasks, is one of many options. As Lanier said in 1988, "[t]here's simply no need for one unified paradigm for experiencing the physical world, and there's no need for one in Virtual Reality either" [Lan88]. This provides great freedom to developers in designing interactions, but inherently makes some design decisions preferable over others. Investigating proposed 3DUI interaction techniques is, therefore, a major topic in HCI and VR research. The choice of design can be evaluated using different methods, such as informal user studies, cognitive walkthroughs, or formal experiments, to gain an initial understanding. For a full evaluation, Bowman and Hodges suggest a fully structured testbed evaluation in which a taxonomy is created, outside factors are considered, and adequate performance metrics are selected to obtain performance measurement, especially in the context of a specific application [BH99]. In IVEs, three important tasks are *navigation*, *manipulation*, and *system control* [LJ+17].

Navigation

Navigating the IVE is equally important for VR as navigating the real environment in the real world. Important tasks related to navigation are the *exploration* of the virtual environment, the *search* for a target location in the form of a *naïve search* (without knowledge about the environment) or *primed search* (with prior knowledge), and the precise *maneuvering* in a local space [LJ+17]. These navigation tasks consist of two components: *Wayfinding*, the cognitive process of identifying the route between source and target location, and *travel*, the physical movement and motor components of action [LJ+17; BKH97].

[5] It should be noted that manufacturer-reported system specifications are often overestimated [Sau+22].

Wayfinding describes the creation and modification of a spatio-cognitive map of the environment [DAA98] in a high-level cognitive task [LJ+17]. For many environments, it is an essential stage in the planning of movement, in which the users decide on the best strategy to reach a target. Wayfinding can be supported by external cues that are encountered in reality in a similar way [LJ+17]. For example, easily visible *landmarks* with a fixed position and orientation within the environment, *maps* showing a 2D aerial representation of the virtual environment, *compasses* that indicate a fixed direction, *signs* that show the direction to a point of interest, *trails* that display previous movements, and *reference objects* with a known size that support the distance estimation. Furthermore, the structural elements of the environment (*paths*, *edges*, *districts*, and *nodes*) enhance the environment's legibility. The planning made in the wayfinding phase determines which skills are utilized in the travel phase of navigation.

Travel describes actual performed movements in the IVE that produce locomotion. According to Bowman, seven key factors determine the usability of a travel technique: 1) how easy it is to use, 2) how easy it is to learn, 3) the spatial awareness during use, 4) its efficacy, 5) the appropriate use of speed, 6) the support of users in gathering information about and from the environment, and 7) the effects on the feeling of presence [BKH97]. Various interaction techniques have been proposed for travel tasks which can be considered natural (e.g., natural walking), semi-natural (e.g., leaning-based or walking-in-place), or non-natural (e.g., using a joystick), depending on their interaction fidelity [Nab+15; MLP16]. Typical approaches for travel in IVEs can be distinguished by their implemented metaphor [LJ+17]:

i) *Walking-based techniques* transfer the movements used in reality as means of locomotion to the IVE. Natural walking describes a 1-to-1 mapping of physical movement in reality to movement in the IVE and can be considered the most obvious choice for movement. Besides the direct mapping of movement, different factors can be applied to change the produced movement in VR. If certain thresholds are not exceeded, the walking speed and rotation can be manipulated in such a way that users do not perceive the manipulation. In these *redirected walking* techniques [Raz05; Ste+09], both the movement speed and the rotation can be increased or decreased by a certain degree without being noticed [Ste+09]. In *scaled walking techniques*, the multiplication factor for movement speed is intentionally higher than this threshold to allow a user to move faster in the virtual environment. In the Seven League Boots technique [IRA07], for example, the speed along the axis of travel is dynamically changed depending on the physical walking speed. Two further types are *walking-in-place* and *human joystick* techniques, which only include some aspects of the movement cycle to trigger movement in a specific direction.

ii) *Steering-based techniques* are traditionally encountered in 3D applications in which the direction of movement is indicated, and specific user actions start and stop the motion, similar to driving a car in reality. In desktop applications, this is often performed using a combination of mouse input for direction and keyboard input for movement at a constant speed. In IVEs, this can also be applied using different means of input, for example, gaze-directed steering using eye tracking or HMD orientation, torso-directed, hand-directed steering using one or more hands, or controller input (e.g., joystick movement or touchpad presses). *Leaning-based techniques* are based on tracking the weight shift of the user's physical body, which is mapped to the movement direction and speed without requiring the user to step in the direction. Steering metaphors can also include physical props, such as a car turning wheel or an airplane cockpit, to simulate a specific type of vehicle. However, in IVEs, steering techniques are often more prone to inducing cybersickness [Uso+99].

iii) *Selection-based techniques* utilize a representation of movement or target. After indicating the movement, it is performed automatically by moving the user along a path (either indicated by the user or calculated by the computer). For example, in a simple setup, the user would first select a target using a pointing ray and would then be moved along a linear path with constant speed to the desired location. A different approach is drawing a path for movement or indicating points on a path for movement that are traveled in succession. An important and popular technique in commercial applications is ray-indicated teleportation [Boz+16], which instantly transports a user to a desired location. Instant travel can, however, reduce spatial orientation, so moving the user along a path or providing other kinds of transitions between the locations can be beneficial in some scenarios [BKH97].

iv) *Manipulation-based techniques* produce movement indirectly through object manipulation. For example, a user can move a camera representation or avatar within a world-in-miniature [SCP95] technique to move the egocentric user's perspective to the indicated position. Single-point and dual-point world manipulation techniques, on the other hand, are direct manipulation-based travel techniques. In these techniques, the user 'grabs' the environment, or a virtual point representing the environment, and pulls or pushes the world according to the intended movement direction.

A large portion of the design possibilities for non-realistic travel techniques (e.g., leaning-based or selection-based techniques) can be covered by a taxonomy proposed by Bowman et al. [BH97] that decomposes the task into three subtasks with different technique components: *direction/target selection*, *velocity/acceleration selection*, and *input conditions* [LJ+17]. By modifying the technique components in this system, it is possible to obtain a related technique and compare the specific effect of a single selected technique component to identify the most efficient design for a specific scenario.

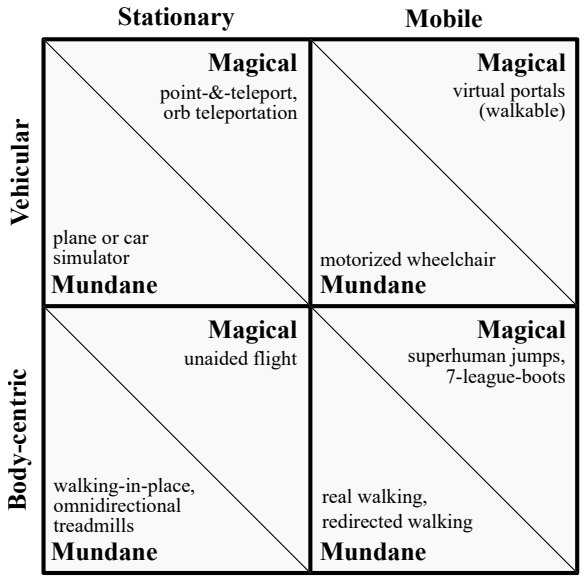


Figure 2.13: A three-dimensional classification system for locomotion techniques proposed by Nilsson et al. [NSN16]. Modified figure adapted from [NSN16].

Nielsson et al. suggest another alternative classification system (see Fig. 2.13) based on the dimensions *plausibility* (a technique adopted from the real world vs. 'magical' techniques), *movement source* (the user's body vs. a utilized vehicle), and *user mobility* (stationary user in a standing or seated position vs. moving) [NSN16], in which technique is localized in a three-dimensional coordinate system. Instead of analyzing the metaphor family or task composition, they analyze how the user moves and how movement is displayed. An advantage of this classification system is the semantic relation between techniques based on their

location within the coordinate system. For example, in a realistic scenario, an adequate locomotion technique could involve a virtual car as a stationary vehicular mundane technique. If realism is not important, a magical device, such as a wand that teleports the user to an indicated location, can be used in a stationary vehicular magical technique.

3D Manipulation

3D Manipulation encompasses the change of spatial properties of rigid three-dimensional objects in the three-dimensional coordinate system of an IVE. The transformation of objects can be described with three position parameters (x,y,z) and three rotation parameters (yaw, pitch, roll), and in many cases, three parameters for scaling along the x-, y-, and z-axis, allowing the manipulation of up to nine degrees of freedom. LaViola et al. suggest four 'canonical' tasks that encompass the fundamental manipulation techniques [LJ+17]:

- *Selection* (also 'target acquisition') is an important action to indicate a specific object or multiple objects to the computer, for example, by ray-casting, virtually touching the object with the hand or a controller, using speech, or image-plane pointing techniques [FHZ96].
- *Positioning* describes the change of the 3D position of an object, typically expressed as a three-dimensional position vector.
- *Rotation* describes the change of the 3D rotation of an object, typically expressed as a three-dimensional rotation vector or four-dimensional quaternion.
- *Scaling* describes the change of the 3D scaling of an object, typically expressed as a three-dimensional scaling vector that contains a scaling factor along each object axis in local space. In contrast to positioning and rotation, this interaction is typically not encountered in reality.

Many manipulation techniques are application-specific and depend on various factors such as object size, distance, the physical and psychological state of the user, and the goals of the application [LJ+17]. For a complete testbed evaluation, Bowman et al. suggest a taxonomy for analyzing manipulation techniques [BH99] In this taxonomy, a manipulation task consists of three subtasks with interaction-specific technique components: First, *selection* with feedback mechanisms, the indication of an object, and the indication to perform the selection. Second, the actual *manipulation*, in which the object is attached to, e.g., the user's hand, to afford the changes regarding position and orientation, and some form of feedback that is provided to the user. In the final *release phase*, the drop event is indicated, and the object is placed at the final location.

Depending on their technique components, manipulation techniques can be considered *isomorphic* when the interaction uses a 1-to-1 mapping to actions in the real world or *nonisomorphic* [LJ+17]. *Nonisomorphic* techniques allow users to manipulate objects in ways that are not possible in the real world. Two one-handed nonisomorphic techniques that incorporate nonisomorphic aspects to allow users to manipulate objects from afar are, for example, HOMER (Hand-Centered Manipulation Extending Ray-Casting) [BH97], which places the hand in the object's vicinity after a ray-based selection, and the Go-Go technique [Pou+96a] that uses a non-linear mapping of the user's arm movements to extend the reach of the user. In both cases, the selection of an object is a nonisomorphic technique component, whereas the actual 3D manipulation is isomorphic. The spindle technique [MM95] employs two controllers that are both moved synchronously and symmetrically. The object is placed at the center point between both controllers to enable positioning. It can furthermore be rotated by moving the hands relative to each other, and also scaled by increasing or decreasing the distance between the two controllers.

System Control

Figure 2.14: An example of utilizing the WIMP concept as an approach to system control in VR: Selection of a menu item using a pointer in the Meta Quest 2 home screen.

The last type of common interaction in IVEs are system control techniques, which are related to specific functions of the system. System control actions correspond to a specific command, such as requesting the execution of a specific function or changing the mode of interaction or state of the system [LJ+17]. In many cases, conventional metaphors, such as a GUI incorporating WIMP concepts (e.g., buttons, sliders, menus, windows, pointers), are utilized in VR applications to exploit the knowledge of users regarding the interaction with desktop computers and traditional 2D screens. In the Meta Quest 2 system, for example, the home environment from which the user starts applications resembles the GUI of a desktop system with typical WIMP elements that can be controlled using hand movements to move a virtual pointer and a pinch gesture to confirm a selection (see Fig. 2.14).

2.2.6 Whole-Hand User Interfaces

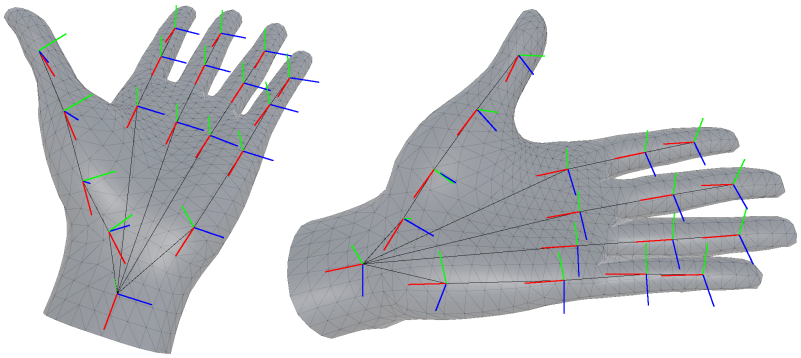


Figure 2.15: The bone hierarchy and joint rotations of the Meta Quest 2 hand model visualized in Unity 3D (Oculus XR plugin 3.0.2).

In his PHD thesis from 1992, Sturman defines whole-hand input as “the full and direct use of the hand’s capabilities for the control of computer-mediated tasks” [Stu92, p. 23] without additional tools such as pens, keyboards, a mouse, or other physical devices. The input is created by changing the pose of fingers, the position of one or both hands, or a combination of these. The human hand is highly dexterous and expressive, with approximately 29 DOF

per hand (23 for joints, 3 for hand translation, and 3 for hand rotation [Stu92; HS14]), and it provides a high degree of proprioception for finger movements [Lon+22]. Due to its flexibility and high number of DoF, the human hand can be used as a universal tool for diverse tasks in diverse application domains, both in reality and in IVEs. According to Sturman, the salient features that make whole-hand input an interesting approach in HCI research are *naturalness*, *adaptability*, and *dexterity* [Stu92]. He describes *naturalness* as referring to exploiting the skills acquired in everyday use of hands in our real environments, which can be intuitively exploited in interfaces. *Adaptability* describes the use of hands for various tasks and easy switching between operating modes. *Dexterity* describes the transformation of complex coordinated movement patterns into dexterous skills that can be employed to perform complex tasks with low cognitive load.

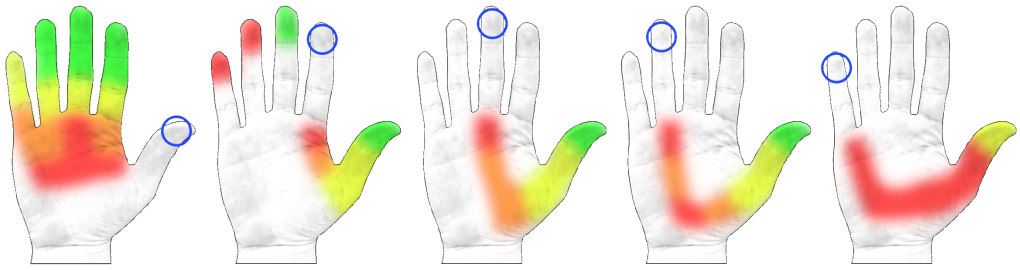


Figure 2.16: Visualization of the functional workspace of the human hand for prehensile gestures for different fingertips as heat maps [DSG19]. The blue ring indicates which tip initiated the movement.

Hands play an important role in both everyday interaction and in HCI [Jer15]. In the real world, humans use their hands in three dominant ways: as *semiotic* tool for communication, as *ergotic* tool for manipulation of the environment, and as *epistemic* tool for tactile exploration [HS14]. The interaction with objects in our environment leads to a large repertoire of subconscious pre-acquired sonsorimotor routines or pre-acquired skills [Stu92], which can easily be employed in new situations. One main use of hands is prehension, the manual handling of objects, including grasping, carrying, manipulating, and releasing them [RA12]. Prehension can be performed as a bidigital (two fingers), pluridigital (multiple fingers), or palmar (including the palm) grip [Kap71]. Six types of prehensile gestures are typically distinguished: *cylindrical grasp*, *tip*, *hook* or *snap*, *palmar*, *spherical* and *lateral* [ST55]. Feix et al. compare 22 publications on grasp types to extract 33 unique prehensile grasp types [Fei+15], which can be considered distinct variants of these base types. Furthermore, prehension can include dynamic components, such as the squeezing of the top of a spray can or eating with chopsticks [Kap71]. The range of prehensile gestures is, for many grasp types, largely determined by the movement range of the thumb [DSG19; Kuo+09] (see Fig. 2.16). The 'pinch'-gesture (visible in Fig. 2.14), a common input action in HCI, can be considered a dynamic bidigital prehensile tip gesture involving the index finger and thumb.

To enable exploiting whole-hand input in HCI, hand movements and finger joint positions can be tracked using diverse tracking systems. These tracking systems yield a structure of virtual bones that is constructed in a similar way to the bone structure in real hands, with bone length and joint orientation corresponding to the user's hands in the real world (see Fig. 2.15). In many cases, this structure is simplified by omitting bones that are not reliably tracked or that cannot be moved sufficiently in reality. For example, the hand model used in the Meta Quest 2 omits three mostly immobile joints of the inner hand (carpal joints), which are considered immobile, whereas the other joints closely match the position and orientation of real-hand bones. Whole-hand input depends on what can be

reliably tracked as well as what can be ergonomically performed by humans. Two primary ways of whole-hand input in VR can be distinguished: gestures and manipulation (see Fig. 2.17).

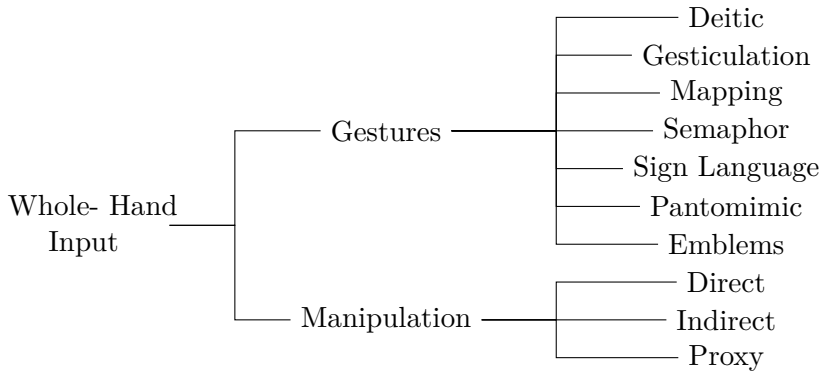


Figure 2.17: A taxonomy of whole-hand input styles.

Gestures Gestural input is a post-WIMP way of interacting with computer systems and a major topic of HCI research. Various forms of input are possible, for example, symbolic, metaphorical, or abstract gestures [WMW09]. Gestures can be performed dynamically or statically, using one or both hands, and with the palm (palmar) or the back of the hand (dorsal) facing the tracking system [Olv+22]. Sturman et al. distinguish between different types of gestures for interaction depending on the mode of implementation (irrelevant, motionless, moving) of the hand and also the fingers (see Table 2.2) [SZP89]. Typically, hand actions and postures in IVEs are exploited as *buttons*, *valuators*, *locators*, and *picks* [SZP89]. The interpretation of hand actions can be direct (physically influencing objects), mapped (movements influence an abstract value), or symbolic (gestures corresponding to a pre-defined interpretation) [Stu92].

Five types of gestures have been described by Karam and Schraefel: *deictic gestures, gesticulation, manipulation, semaphores, and sign language* [KS05]: *Deictic gestures* are used in both real-world and computer interaction to specify a location or direction, such as pointing with a finger at an object. This natural behavior in communication is transferred from the real world to computer systems and can also be expanded by visually enhancing the indicated object or the direction of pointing. In the same way, the natural communication of humans is exploited by implementing *gesticulation* in interfaces. *Gesticulation* or *coverbal gestures* [BH92] accompany our everyday speech, often in a subconscious way, without being taught in an idiosyncratic way [Wex95]. *Manipulation*[6] refers to the mapping of physical movement of hands or fingers to the manipulation of an entity, such as an object or an abstract value. For example, lifting the hand to increase the loudness of a speaker system by mapping the volume to the user's hand height above the ground, or tracking the distance between landmarks, especially the fingertips, and mapping the value changes to arbitrary values. *Semaphors* are small, dynamic, or static abstract signaling gestures that contain no inherent meaning. Similar to a dictionary, a *semaphore* translates to a specific system command. *Sign languages*, on the other hand, are based on sequences of linguistic symbols and communicate meaning to a computer system by forming dialogue-like sentences. Individual symbols, such as letters in a sign language alphabet, can be seen as semaphores, whereas the actual command, which is communicated to the computer system, is formed by a semantically and syntactically correct sequence of

[6] In the taxonomy in Figure 2.17 referred to as 'Mapping' to avoid confusion with the group 'Manipulation.'

symbols in a similar way to text entry in a CLI. Aigner et al. describe two more types of gestures [Aig+12]: *pantomimic gestures*, which visualize task-related actions, and *iconic* gestures, which communicate shapes, sizes, and paths. Another category is *emblems*, which encompasses culture-specific signs, such as "thumbs up" or "V for victory" [HS14].

HAND POSITION & ORIENTATION	FINGER FLEX ANGLES		
	not relevant	**motionless fingers**	**moving fingers**
not relevant	X	finger posture	finger gesture
motionless hand	hand posture	oriented posture	oriented gesture
moving hand	hand gesture	moving posture	moving gesture

Table 2.2: Taxonomy for gestures and poses of hand and fingers. Adapted from [SZP89].

Manipulation In the taxonomy in Figure 2.17, the term 'manipulation' is used to refer to the physical handling of objects in IVEs resembling physical interaction with objects in reality. A direct manipulation that is similar to handling a real-world object, also referred to as "virtual hand" [HS14, p. 299], allows for high performance in manipulation tasks [MBJS97]. Requirements for high-fidelity direct manipulation are accuracy regarding hand tracking, fast response to user actions, and adequate feedback [Bry05]. A high-fidelity interaction exploiting the fine articulation of the human hand remains a challenging task with today's consumer technology, due to the limited tracking capabilities and missing haptic feedback. Therefore, this form of interaction is often abstracted and does not faithfully replicate real-world manipulation of objects. For example, a medium-fidelity direct manipulation of virtual objects using hands could be initiated by a selection event, e.g. a pinching posture with thumb and index finger performed in close vicinity to the object, followed by a manipulation phase, in which the translation and rotation of the real hand are coupled to the transformation of the virtual object, and, finally, terminated by releasing the pinch posture [Buc+04; JFH94].

Indirect manipulation, on the other hand, refers to the manipulation of objects that influences specific properties of other objects [Bry05], for example, changing the position of a virtual slider through direct manipulation to change the brightness of a virtual light bulb. In proxy manipulation [HS14], a proxy object represents another object, and 3D transformations of the proxy object are mapped to the represented object [Jer15]. For example, in a world-in-miniature technique [SCP95], a miniature version of a virtual furniture piece can be placed within a miniature version of the virtual room via direct manipulation, which is then transferred to the full-size virtual object in the full-scale room to enable the manipulation of big objects from afar with less physical effort.

2.3 Human Factors

The term "human factors" refers to aspects (limitations, capabilities, working principles) of the human body, senses, and thinking that need to be addressed in the design of interactive systems [LJ+17]. In this section, four theoretical approaches are presented that form the scaffolding of the conceptual part of this thesis:

- *Schema theory*, a fundamental view on the structure of knowledge (section 2.3.1).
- *Information processing theory* as a fundamental approach to cognition (section 2.3.2).
- *Enactivism*, an alternative approach to describing cognition (section 2.3.3).
- *Activity theory*, a social perspective on acting within a community (section 2.3.4).

2.3.1 Schema Theory

> "A schema theory is basically a theory about knowledge. It is a theory about how knowledge is represented and about how that representation facilitates the use of the knowledge in particular ways. According to schema theories, all knowledge is packaged into units. These units are the schemata. Embedded in these packets of knowledge is, in addition to the knowledge itself, information about how this knowledge is to be used." [Rum17, p. 34]

Units of Knowledge Schemata[7] (Greek for 'image') are considered a fundamental mental construct in cognitive science that organizes ways of thinking and acting by providing templates for understanding and acting upon the world [Rum17; NS86; Arb92; Swe03; Ber+08]. These units are constructed by users from experiences of acting within the world and provide the building blocks of intellectual development. In this context, the term 'schema' is used to describe diverse types of knowledge, such as behavior, mental activities, and mental symbols [Ber+08], cognitive and abstract procedures [Rum17; Swe03], and implicit and subconscious knowledge in the form of the *body schema* [Gal86], the schematic representation of *motor skills* [New91], and sensorimotor schemata [DPBB17] in enactivism (see section 2.3.3). According to Sweller, learning can, in conscious cognitive processes, be considered a manipulation of existing schemata that results in the "creation of new, higher order schemas and automation" [Swe03, p. 222]. Schema acquisition and automation are necessary to handle information-rich material, and our cognitive architecture "has evolved so that very high element interactivity material encompassing large amounts of information can *only* be handled when incorporated in schemas" [Swe03, p. 224]. Rumelhart describes six general properties of schemata [Rum17]:

1. Schemata include variables that are filled with information during an interaction. This enables a flexible application of schemata to account for diverse situations.
2. Schemata are embedded in super-schemata and embed sub-schemata themselves [Chi+81], which makes them hierarchically structured. At the lowest level, schemata describe elementary primitives that are combined to form complex structures.
3. The knowledge in schemata is represented at all levels of abstraction. It describes the nature of events, objects, and situations in an informal, private, and unarticulated way.
4. Schemata represent generalized knowledge and generic concepts rather than strict definitions.
5. Schemata are active processes that carry out their own tasks, such as perceiving, learning, understanding, and solving problems.
6. One of these active processes is the recognition of the goodness of their fit.

Based on a combination of the results of various studies, Gilhooly lists five *expertise maxims* that describe the benefits of successfully encoding complex information and procedures in the form of schemata and applying these to solve problems [Gil90]. According to these maxims, experts are (1) more efficient and rapid at encoding problem details in memory. (2) They work forward and directly apply schemata to solve problems (*schema-driven*) instead of searching for solutions in a means-end analysis (*search-driven*). (3) The level of understanding and representing problems is, in experts, deeper and more sophisticated, and (4) their overall knowledge in the field is better, which facilitates identifying and solving

[7] Besides 'schemata,' 'schemes,' 'schemas,' and 'schematas' are also common plural forms. In this thesis, 'schema' and 'schemata' as a plural form are used, except for direct quotations.

familiar problems. Furthermore, (5) expert knowledge and the associated schemata have been acquired through extensive practice.

Constructivism The term schema appears in the constructivist theories of Jean Piaget, who analyzed the cognitive development of children. He proposed that children construct their understanding of the world through their experiences and interactions, forming schemata as mental templates for interaction. Whenever a new experience does not fit into the existing structures, the internal balance, the *equilibrium*, is disturbed. In Piaget's view, this disturbance can be reduced by either *accommodating* the model to new experiences or by *assimilating* new experiences into existing structures [Pia52]. A key element in Piaget's theory is that schemata are, in their fundamental form, based on experiencing simple physical interactions with the world as a child. With maturing and accumulating knowledge about the world, they form increasingly complex structures that represent complex and abstracted knowledge required for mathematical concepts, ethics, language, and culture [Kim05; Lak12]. This knowledge forms a coordinated and interconnected network of schemata, in which every schema "constitutes a totality with differentiated parts. Every act of intelligence presupposes a system of mutual implications and interconnected meanings" [Pia52].

> "Whereas the initial schemata are only interconnected due to their reflex and organic substructure, the more evolved schemata, at first primary, then secondary and tertiary, become organized little by little into coherent systems due to a process of mutual assimilation which we have often emphasized and which we have compared to the increasing implication of concepts and relationships. Now, not only is this progress of assimilation correlative to that of accommodation, but also it makes possible the gradual objectification of intelligence itself." [Pia52, p. 413]

According to Fischer, skills and cognitive processes begin as sensorimotor schemata (in [Fis80] called *sensory-motor actions*) that contain both actions and perceptions. These schemata are "purely practical: [Infants] understand how to act on specific things in the world but cannot think about those things independently of acting on them." From these schemata, so-called *mappings* are constructed in which two schemata are coordinated. From *mappings*, *systems* emerge, and finally, *systems of systems*. These basic *systems of systems* form a *representation*, which is ultimately further developed through combinations with other representations to form *abstractions* [Fis80]. Similarly, Derry proposes different categories of *memory objects* with different levels of complexity [Der96]: On the lowest level, *phenomenological primitives (p-primes)* encompass basic and intuitive abstractions of common events. Well-structured *p-primes* form complex schemata that enable, for example, pattern recognition and arithmetic expressions. At the highest level, several schemata form an *object family* that associates related schemata within a single domain. A collection of diverse *memory objects* forms the *mental model*, which is employed to understand a phenomenon and form a foundation for reasoning and problem-solving. Finally, *cognitive fields* mediate between experience and *memory objects* by triggering the activation of schema structures, making them more available to use than others for a specific problem or in a corresponding situation.

An advantage of using schemata to solve problems is that, by integrating a large amount of complex knowledge into a schema stored in long-term memory, the limitations of working memory are much less relevant for performing a task [Swe03]. Sweller suggests, among other continua, a *cognitive matrix of continua* to describe the different requirements for new material and well-learned material (see Table 2.3) [Swe03]. In this matrix, new information is contrasted with familiar information. For new information, no schemata exist that

		Processing new information		Processing familiar information
1.	Learning continuum	New material	—	Well-learned material
2.	Central executive function continuum	No schema-based central executive function	—	Schema-based central executive function
3.	Problem-solving search continuum	Problem-solving search required	—	No problem-solving search required
4.	Element combinations continuum	Combinations of elements random	—	Combinations of elements ordered
5.	Working memory limitations continuum	Working memory limitations relevant	—	Working memory limitations irrelevant

Table 2.3: Cognitive matrix of continua proposed by Sweller [Swe03].

encode the information in an efficient way, which makes a problem-solving search necessary, and limitations of working memory apply. For familiar information, on the other hand, schemata exist that serve as a central coordinating executive function, which efficiently encodes the information and enables a direct application of skills.

Image Scheme Theory In cognitive linguistics, the concept of image schemata as abstractions of everyday experiences is discussed. Many expressions in our language are figurative, and the focus of research lies in analyzing the relationship between real-world experience and abstract and representational thought. Johnson defined an image schema as "a recurring, dynamic pattern of our perceptual interactions and motor programs that gives coherence and structure to our experience" [Joh87, p. xiv]. Hampe describes image schemas as flexible and internally structured experiences that integrate recurrent sensory-motor and spatial experiences, as well as body movements, into meaningful, pre-conceptual structures [Ham05]. How we perceive objects and the world is inseparably connected to the (physical) ways "in which humans habitually use or interact with those objects" [RL+78].

Lakoff and Johnson have proposed the concept of image schemata, or so-called 'cogs' (*cognitive primitives*) [Lak12] as a key element in processing complex and abstract information. Lakoff et al. "hypothesize that primitive concepts have a schema structure that mediates between embodiment circuitry and complex concepts that are expressed by linguistic structures in natural language" [Lak14]. Metaphors and linguistic constructs are, according to this theory, grounded in the sensorimotor experience of life and form certain metaphorical concepts, such as movement ('working towards a goal'), container ('the heart containing a feeling'), and spatial perception ('feeling up or down'), which can be described as an abstract representation of experiencing a physical existing. The group of *aspect expressions*, for example, contains concepts about progression (*enable, ready, cancel, start, process, suspend, fail, finish, iterate* and *result*) [Nar97] and can be considered a fundamental aspect of perceiving and controlling tasks. The acquired image schemata structure the human system of thought [Lak12]. At the very core of image schemata, as proposed by Johnson and Lakoff, lies the assumption that "the primary tools for thinking are metaphors, as opposed to logical inferences, and it is through reframing problems and situations in an embodied context that we normally find their solutions" [DPBB17]. Many of these metaphors can be considered to be constructed from *repeated sensorimotor occurrence*, forming *primary metaphors* which are then *instantiated* as a general principle in the form of language, behavior, and thinking [Hur11]. Based on [Joh87], Hurtienne and Israel provide a grouping of common image schemata as conceptual ideas about action (see Table 2.4).

Group	Image Shemata
Basic schemas	SUBSTANCE, OBJECT
Space	UP-DOWN, LEFT-RIGHT, NEAR-FAR, FRONT-BACK, CENTER-PERIPHERY, STRAIGHT-CURVED, CONTACT, PATH, SCALE, LOCATION
Containment	CONTAINER, IN-OUT, CONTENT, FULL-EMPTY, SURFACE
Identity	FACE, MATCHING
Multiplicity	MERGING, COLLECTION, SPLITTING, PART-WHOLE, COUNT-MASS, LINKAGE
Process	SUPERIMPOSITION, INTERACTION, CYCLE
Force	DIVERSION, COUNTERFORCE, RESTRAINT REMOVAL, RESISTANCE, ATTRACTION, COMPULSION, BLOCKAGE, BALANCE, MOMENTUM, ENABLEMENT
Attribute	HEAVY-LIGHT, DARK-BRIGHT, BIG-SMALL, WARM-COLD, STRONG-WEAK

Table 2.4: Categorization of image schemata as proposed by Hurtienne and Israel [HI07].

According to Mandler and Cánovas, the emerging cognitive structures can be described at three levels [MC14]: (1) *Spatial primitives*, which are grounded in spatio-temporal interaction, (2) *image schemas*, which combine multiple spatial primitives, and (3) *schematic integration*, which includes non-spatial elements (e.g., feelings) into schemata. Spatial primitives are directly derived from early physical interactions with the world. For example, at some point in their cognitive development, infants are confronted with physical objects that possess an IN-side and an OUT-side, a CONTAINER. Objects can be placed with-IN the CONTAINER, which combines the spatial primitives CONTAINER, OBJECT, and IN into the image schemata OBJECT-IN-A-CONTAINER [MC14]. This image schema can be abstracted to enable a metaphorical description and becomes, in this process, a *schematic integration*. For example, the abstract process of thinking about something is expressed as "having something in one's mind."

Schemata in HCI The concept of schemata as knowledge representations is widely accepted as a unit to describe the underlying structure and processes of action and cognition in psychology [KM05; Chi+81]. In comparison to other more established methodologies, the amount of research literature regarding the direct analysis of the content of applied schemata in HCI is, however, limited [Car03], making it more of a niche topic in VR and HCI research. Schemata have been used to describe the knowledge about interactive elements in interfaces. Rohrer, for example, describes that *feeling-based user interfaces*, as opposed to *conversation-based user interfaces*, require "thinking about subjective, preverbal bodily patterns of feeling ... instead of abstract, symbolic, and verbal communication" [Roh95], which corresponds to the idea of image schemata. Hurtienne describes the intuitive use of interface elements if they are based on corresponding image schemata and proposes that image schemata form an expressive language for analyzing user interfaces and describing phenomena [Hur11; Hur17]. Jetter et al. integrate the research of Hurtienne et al. in their *blended interaction* framework and advocate for the conceptual integration of image schemata, Jacob's themes of reality [Jac+08b], and affordances into the interaction design to offer users a natural interaction with "the unfamiliar expressive power and "magic" of the digital world" [JRG14]. Hedblom analyzes the recurring patterns of actions in interfaces based on image schemata and introduces a formalized *image schema logic* to express relationships [Hed19]. Antle et al. describe, in reference to image schema theory, the use of schematic *embodied conceptual metaphors* as a linkage between an embodied

schemata as source and an abstract concept as target domain, allowing users an intuitive interaction with technology [ACD09]. In VR interaction, image schemata are commonly encountered in spatial interaction that mimics the interaction in physical reality [MB05]. In a similar way, Richir et al. describe the "imported behavioural schema" that mimics the real world and enables an intuitive *pseudo-natural* interaction in IVEs by exploiting and adapting real-world knowledge [Ric+15].

2.3.2 Information Processing Theory

An approach that has been very influential in describing the working principles of human cognition is information processing theory, with its roots in psychological research in the 20th century. This cognitivist view follows the reductionist idea that cognition and behavior are the result of transmitting and processing information in sub-components (or modules) in a central processing unit (the brain). In this model, cognition can be completely described as a brain-centered process based on modifying mental representations and symbols representing the outside world or abstract problems [Ber+08]. This theory does not propose that all human cognition and behavior can be modeled in this way, but certain domains, such as HCI, can be effectively modeled in this framework that allows for analyzing tasks and predicting performance [Car03]. Winograd and Flores summarize the key assumptions of cognitivism and information processing theory as follows:

> "1. All cognitive systems are symbol systems. They achieve their intelligence by symbolizing external and internal situations and events, and by manipulating those symbols.
> 2. All cognitive systems share a basic underlying set of symbol-manipulating processes.
> 3. A theory of cognition can be couched as a program in an appropriate symbolic formalism such that the program when run in the appropriate environment will produce the observed behavior." [WF86, p. 25]

Based on this general assumption of information processing theory, Card et al. proposed the Model Human Processor (MHP) [CNM83] to describe cognition in the context of performing tasks in HCI in a direct analogy to how computer programs perform calculations based on the provided input. It offers both a description of the flow of information and further experimental estimations of timings and capacities for the individual modules and flows. The MHP is divided into three distinct main modules (depicted in Fig. 2.18) that fulfill a specific step in transforming the input stimulus from the environment to cognitive processes and ultimately to a physical response: the *perceptive processor,* the *cognitive processor*, and the *motor processor*. Furthermore, the two resources *attention* and *memory* influence the functioning of the modules and processing of data. These central components are briefly described in the following paragraphs.

The Perceptual Processor As the first step in the information processing pipeline, the information stimuli are gathered through the sensory system. Each sensation relies on specialized mechanical, photoreceptive, or chemical sensors [Sch+95] that receive a stimulus in the form of energy, e.g., mechanical pressure and electromagnetic waves from the environment, that is encoded as a neural activity signal and transmitted to the brain via the nervous system [GC21]. The incoming neural activations are processed in the brain and produce perceptions, which are further processed in the cognitive pipeline. The process of combining different stimuli into a coherent object in the conscious perception is also referred to as "binding" [Jer15, p. 72].

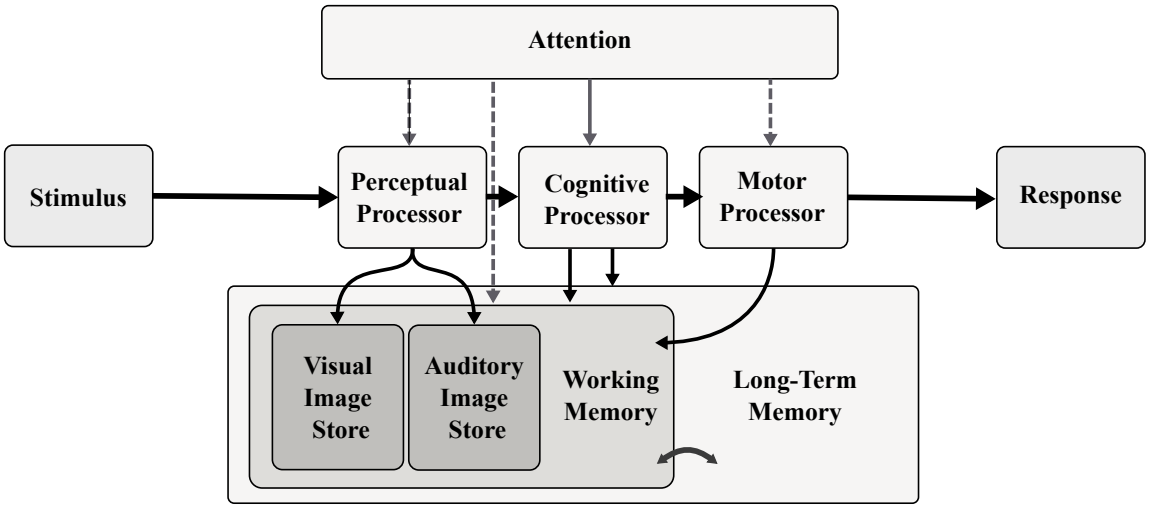


Figure 2.18: The modules of the cognitive pipeline described in the model human processor [CNM83].

Sensory perception can be divided into two categories: *exteroceptive* (perceiving the environment) or *interoceptive* (perceiving one's own body). For exteroception, the five main senses to perceive our environment are visual, auditory, tactile (pressure & temperature), gustatory, and olfactory perception. The most important sensory information of these are haptic (tactile), echoic (auditory), and iconic (visual) information [CG17], which are also the senses targeted in VR applications. For interoception, the specialized sensory systems are located inside the body to provide, for example, vestibular (balance and spatial orientation of the head) and proprioceptive (detecting the spatial location of body parts) perception.

The Cognitive Processor The cognitive processor is the central sub-component that transforms all incoming stimuli, as well as physiological needs, motivations, and prior experiences, into an output, typically a motor response, such as typing on a keyboard or using voice. Two intertwined modes of operating can be distinguished in human cognition: *action-selection* and *decision-making* [LJ+17]. Action-selection describes the automatic response to inputs based on some form of simple cognition and association, whereas decision-making encompasses a more sophisticated mode of thinking and complex processing of information [LJ+17].

This concept has been formulated as the dual system (or dual process) theory. Dual system theory proposes that human cognition consists of two quasi-distinct systems for knowledge and cognition, the automatic and intuitive *System 1*, and the reflective *System 2*, which is used for reflective thinking and hypothetical reasoning [KFD15; Eva03; Shl12] (also called Type 1 and Type 2 by some authors). In this dual-process theory, System 1 is a primordial system of thinking that is shared with animals [Eva03]. It produces "rapid, intuitive, automatic response" [Shl12] and can be characterized as "rooted in memory and rehearsed experience, more given to analogy than to logic" [KFD15]. System 2, on the other hand, is "conscious, slow, controlled, deliberate, effortful, statistical, suspicious, and lazy (costly to use)," [Shl12] and "concerned with novelty or perceived incoherences" [KFD15]. While both systems are separated, the strategy employed to solve tasks can be grounded in both, and sufficient exposure to a number of cases leads to shifts from System 2 to System 1. System 2 is used for new and unfamiliar information, whereas System 1 is activated for well-known material with a "schema-based central executive function" [Swe03]. Skills that shift from System 2 to System 1 become automated and require less conscious effort and,

expressed in the model of information processing, less working memory load [Swe03]. Only in cases where an adequate solution cannot be obtained using System 1, the more costly System 2 is employed to find new solutions [KFD15] in a process of deliberate thinking. Systems 1 and 2 can be considered two modes of operation of the cognitive processor that depend on the familiarity of the encountered problem.

Human decision-making, which corresponds to System 2 thinking, can be based on different strategies (see, for example, [KS14; Ber+08]), such as:

- *Heuristics*: Applying simple rules that work well in most circumstances.
- *Fast-and-frugal strategies*: Making decisions based on the most important cues in the environment.
- *Prospect theory*: Selecting the most promising prospective outcome in comparison to a subjective reference point.
- *Support theory*: Combining diverse beliefs and likelihoods from multiple sources to identify the most justifiable strategy.
- *Means-end analysis*: Deciding the next steps based on a decomposition of goals into manageable sub-goals.
- *Working-backward strategy*: Selecting the intended outcome of decisions and tracing the necessary steps back to the starting point.
- *Analogies*: Solving new problems by identifying and transferring strategies from previously encountered similar situations.

Cognition is constrained by limited resources, and the steps required for any mental activity, such as transferring knowledge from long-term memory to short-term memory or dividing attention between multiple tasks, increase the *cognitive load*, which, as a result, reduces task performance [Swe03; Hol+10]. Three types of additive cognitive load are distinguished when users encounter new material: *intrinsic cognitive load* (based on the complexity of information), *extraneous cognitive load* (inappropriate representation of information), and *germane cognitive load* (as a result of knowledge construction that contributes to learning) [Hol+10]. In contrast to novel problems that require active processing of information, familiar problems can be solved with much less cognitive load by applying acquired schemata [Swe03].

The Motor Processor As the final step in the information processing pipeline, the results of information processing in the cognitive processor activate patterns of muscle activity as output or system response to input. These activation patterns trigger muscle groups, often opposing agonist and antagonist muscles, that control actions. In the case of traditional desktop computers, these primarily include head and eye movements that change perception and arm and finger movements that provide input to the computer system [CNM83]. In full-body 3DUIs, however, this includes all parts of the body, for example, leg and foot movements, as well as twisting the torso. The input information provided to the motor processor by the cognitive processor can have different functions: as *prescription*, the construction of memory constructs required for skill acquisition, as *feedback* about the results of actions and the current ongoing action, and for *search-channeling*, to locate task-relevant muscle activation patterns within the perceptual-motor workspace [New91].

The neurons that encode muscle activation for a distinct part of the human body are located in specific zones of the brain in the motor cortex. Similarly, tactile perceptions of specific body parts are located in the somatosensory cortex [PB37; PR50; Rou+20]. The portions of the brain that are associated with specific parts of the body can be visualized as

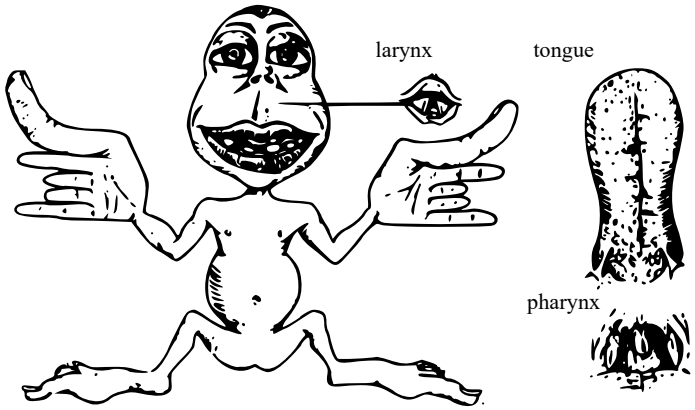


Figure 2.19: The original homunculus (assembled) including the tongue, larynx (for vocalization), and pharynx (for swallowing) proposed by Penfield and Boldrey. Limb proportions correspond to the associated portion in the brain tissue [PB37].

Figure adapted from [PB37].

the cortical homunculus (see Fig. 2.19) [PB37; PR50]. This figure visualizes how specific body parts that are responsible for perceiving the outside world and performing actions, such as thumbs, lips, tongue, eyes, and toes, are represented by a larger portion of the brain in relation to their size in comparison to the human body.

Memory The processing of information is dependent on previously acquired knowledge about the world and interactions that are stored in memory. Two types of knowledge can be distinguished: declarative (explicit) and procedural (implicit) knowledge [EHK18; LJ+17]. Declarative knowledge encompasses factual knowledge that can be easily expressed and is characterized by conscious access to information [DE10]. It encompasses semantic memory and episodic memory. *Episodic memory* stores individual experiences, their time and place, and detailed information on the event [DE10], whereas *semantic memory* contains all individually acquired factual and believed knowledge about the world. Procedural memory, on the other hand, is characterized by unconscious access to information and implicit changes in behavior and responses to stimuli [DE10]. According to Carmina and Güell, implicit memory can be divided into four sub-types [CG17]:

1. *Procedural memory (habits and skills)*: Subconscious retrieval and automatic execution of motor and executive skills essential for task performance, often trained over a period of time.
2. *Associative memory (classical and operant Conditioning)*: A stimulus can be coupled with other stimuli to trigger a response or behavior. In classical conditioning, the response to an unconditioned stimulus can be associated with a different stimulus if both are present at the same time. Operant conditioning describes the acquisition of new behavior by acting and associating perceived consequences with the behavior.
3. *Non-associative Memory (habituation and sensitization)*: Habituation describes a decreased response to a repeatedly perceived stimulus, whereas sensitization describes an increased response.
4. *Priming*: A previously perceived stimulus influences the response to a later perceived stimulus.

An influential model that describes the access and retrieval of knowledge is the Atkinson-Shiffrin-model which describes the human memory at three different layers with distinct functions and control processes: The sensory register (in Fig. 2.18 called *Auditory Image*

Store and *Visual Image Store*, and embedded into working memory), working memory (also referred to as short-term memory), and long-term memory [AS68]. The first layer receives the perceived sensory information and stores information for a short amount of time, which makes perceptions immediately available for cognitive processing. The short-term memory contains a limited amount of heterogeneous information that is important in the context of interaction. Information is stored in the form of chunks, which group information [Ber+08] to reduce the required capacity. Humans can maintain 7 +/- 2 of these chunks in their short-term memory, often referred to as the *magic number* [Mil75], with the first and last items of a sequence generally better retrievable in serial recall tasks [MJ62]. Information stored in the working memory is typically stored for several seconds and either transferred to long-term memory or forgotten [Ber+08]. In long-term memory, on the other hand, a large amount of information can be stored for a long time, and it decays only slowly when not retrieved [AS68].

The acquisition of skills is separated into three phases [FP67; CG17]: i) the cognitive phase, in which skills are actively deconstructed into understandable sub-components, and the correct execution requires attention; ii) the associative phase, in which an increasingly automatic response pattern is created from repeated practice; and, finally, iii) the autonomous phase in which only low cognitive processing is required to respond to stimuli with the correct pattern.

Attention Besides memory, attention is a limited resource that influences cognitive processes by selecting information for further processing and inhibiting other information from being processed [KS14]. In any given situation, attention supports cognitive processes by deciding which information is most relevant to producing an adequate response. Attention can be divided into three types:

- *Selective attention*: Choosing which processes or events receive attention.
- *Focused attention*: Concentrating the attention resources on one process and excluding others.
- *Divided attention*: Parallel processing of multiple processes at a given point in time.

Attention can work as a *bottom-up* process that is driven by sensory information, for example, one element with a different shape in a spatial arrangement of otherwise identical elements captures our attention automatically (also called *visual salience* [GC21]), or it can be controlled *top-down*, for example, in the search of an element with specified properties in a group of different objects [KS14]. In a top-down scenario, the task determines how attention is shifted [GC21], for example, Yarbus found in an experiment that the eye movements of participants depend on the specific task of analyzing an image [Yar67].

2.3.3 Enactivism

The Enactive Approach Enactivism is a theoretical framework that challenges traditional cognitive science's view of the mind as a passive, causally-closed [Ell+14] computational information-processing system that is (in the sense of the Cartesian dualism) separated from its body and environment. In this regard, the information-processing paradigm has been criticized for being "biased [...] toward the disembodied and intellectualist end of the spectrum, the kind of explanations that, to many people, do not match well the situated and richly context-dependent experiences and activities they enact every day" [DPBB17]. The enactive approach emphasizes the inseparable and co-constitutive relationship between perception and embodied action in the shaping of cognitive processes [Col20]. Cognition and the construction of meaning are distributed throughout the agent-environment system, and knowledge is seen as being grounded in embodied action [Hol10]. The two fundamental

claims of enactivism are: "(1) perception consists in perceptually guided action and (2) cognitive structures emerge from the recurrent sensorimotor patterns that enable action to be perceptually guided" [VTR93]. Two important concepts in enactivism are *autopoiesis*, which describes cognition as a biological process of an autonomous, self-referencing, and self-constructing entity [VTR93; MV87], and *co-determination* of agents and the world, in which the domain of problems and meaning are constructed from the properties of the agent and the world [VTR93].

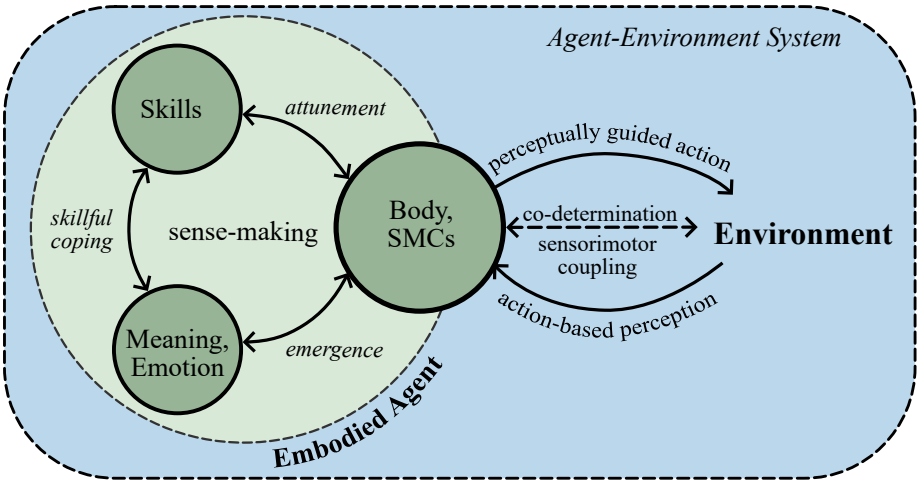


Figure 2.20: Visualization of major key elements of enactivism. Reprint from [DGS23].

The Agent-Environment System Di Paolo defines an agent "as an autonomous system capable of adaptively regulating its coupling with the environment according to the norms established by its own viability conditions" [DPBB17] with adaptivity described as "the capability of an autonomous system to respond to tendencies in the trajectories of its states and its relations to the world, so that when these tendencies approach the boundary of its own viability the system modulates its coupling with the world in a way that tends to avert the crossing of this boundary" [DPBB17]. The agent possesses agency within an environment if the agent shows (1) *autonomy* so that the processes of the agent are interdependent and the agent can be distinguished from the environment, and (2) if the agent shows *adaptivity* by modulating properties and processes to maintain its functioning and existence within the environment [DPBB17]. Every agent exists in a constant active struggle to maintain its organization using its physical features in accordance with the properties of the environment (*sensorimotor coupling*), which requires an adaptation to and structuring of the environment [DP05]. The continuing process of autonomously adapting to changes in the environment to provide significance and valence to the agent [Tho07] leads to the emergence of an agent-environment-specific 'cognition' and agent-specific 'meaning', which can be characterized as "the world invested with interest for the agent itself" [DPBB17].

Agents continuously influence their environment and regulate their structural coupling with it. They actively develop their understanding of the world through their embodied interactions in a process called 'sense-making.' Sense-making relies on the action-perception loop, specifically, on perceptively guided action and action-based perception during engagement with the environment, and encompasses the "capacity of an autonomous system to adaptively regulate its operation and its relation to the environment depending on the virtual consequences for its own viability as a form of life" [DPCDJ18]. The dynamic interplay between perception and action forms the foundation of cognition and behavior, which aims to ensure the continuity of the self-generated organization, establishing a world- and agent-

dependent normative perspective on action [DPRDJ10]. Cognition, as well as affective and emotional responses to the environment and to actions [Col14], emerge from the dynamic interaction between a physical agent and its environment [WSV17]. “To ”make sense” is to relate, to complete, to coordinate one thing with another so that ”sense” or understanding arises. The enactive knower is active in making sense of its environment as it creates its life” [RS20].

For enactivists, the prime example of an agent as a cognitive system is a single-cell organism that, regardless of not possessing a central brain as a processing unit, is able to show quasi-intelligent behavior in performing chemotaxis, the navigation within its environment to search for food using specialized chemical receptors and mechanical effectors. This simple form of cognition, sometimes referred to as *basic mind*, is the foundation of life and is present in the simplest organisms. Cognition and subsequent actions are always related to some intrinsic norms of the organism in the organism-environment system [Tho07]. The purpose of cognition in an autopoietic organism is to enable it to maintain its structural integrity, adapt to changes in the environment, and autonomously perform meaningful actions. All agents and living organisms are entities relying on some form of cognition, constituted by their physical features and properties of the environmental surroundings. Both “life and mind share a set of basic organizational properties, and the organizational properties distinctive of mind are an enriched version of those fundamental to life” [Tho07]. A general definition in the framework of enactivism for a cognitive system based on the idea of the continuity of life and mind is provided by Maturana:

> “A cognitive system is a system whose organization defines a domain of interactions in which it can act with relevance to the maintenance of itself, and the process of cognition is the actual (inductive) acting or behaving in this domain. *Living systems are cognitive systems, and living as a process is a process of cognition.*” [Mat70, p. 13]

Sensorimotor Contingecies The physicality of existence and the physical coupling between an agent and its environment are of major interest in enactivism. The different potentials for physically engaging with the environment, the *sensorimotor coupling*, are called *sensorimotor contingencies* (SMC)s. Several definitions exist that characterize these SMCs [DPBB17]. Di Paolo describes SMCs as “co-variations of sensory stimulation, neural, and motor activity” [DPBB17], and distinguishes four stages of SMCs: (i) First, the *sensorimotor environment* encompasses “the set of all possible sensory dependencies on motor states ... for a particular type of agent and a particular environment” [DPBB17]. On this first stage, actions and perception are treated as open-loop systems that are considered independent from sensory feedback [DPBB17]. (ii) The second stage, the sonsorimotor habitat, describes the “set of all sensorimotor trajectories (i.e., movements in sensorimotor space) that can be generated by the closed-loop system in a given situation” [DPBB17]. (iii) Coordinated patterns of closed-loop systems form “a *sensorimotor coordination* [emphasis in original] if it contributes functionally to the performance or goals of the agent” [DPBB17]. They are “specific, often local sensorimotor co-dependencies that are dynamically organized in time in the context of a task” [DPBB17]. (iv) The finale stage, the *sensorimotor scheme* “describes an organization of [sensorimotor] coordination patterns that is regularly used by the agent because it has been evaluated as preferable (along some relevant normative framework) for achieving a particular goal” [DPBB17]. The normative framework, according to Di Paolo, “involves reference to given criteria that distinguish or value some possible outcomes as preferable to others” [DPBB17].

SMCs are always dependent on both the physical body of the agent and the properties of the environment. Developing and fine-tuning SMCs leads to a state of 'attunement' in which a set of specific skills is developed that enables the agent to pursue meanings. These skills can actively be employed to skillfully cope [Dre14] with the challenges of the environment by selecting and modulating SMCs according to situated and normative factors [DPBB17]. The correct selection and modulation of SMCs depending on the environmental circumstances can also be described as a 'skillful mastery' of the laws of the agent-environment-specific SMCs [DPBB17], in which the agent shows "a regulated openness to be coupled to the world and to be guided by it starting from what has worked in the past" [DPBB17]. SMCs determine how agents interact with their world and ultimately allow for a specific perception and subjective experience of reality and, subsequently, the emergent specific type of cognition shown by an agent in the form of the "regulated sensorimotor coupling between a cognitive agent and its environment" [FDP11]. A visual example used in enactivism to illustrate the sensorimotor coupling between an arbitrary agent and its environment and the subsequent form of physicality-based cognition originated in the philosophy of Merleau-Ponty, the blind man using a walking cane for navigation:

> "Think of a blind person tap-tapping his or her way around a cluttered space, perceiving that space by touch, not all at once, but through time, by skillful probing and movement. This is, or at least ought to be, our paradigm of what perceiving is. ... [A]ll perception is touch-like in this way: perceptual experience acquires content thanks to our possession of bodily skills. What we perceive is determined by what we do (or what we know how to do); it is determined by what we are ready to do ..., we enact our perceptual experience; we act it out." [Noë04]

The task of navigating an environment is enabled through physical interactions that are transformed into subjective perceptions and gestalts from which meaning and cognition emerge for the agent. Cognition and higher cognitive functions, such as deliberation, reflection, and imagination, have to be seen as a dynamic process that is distributed across the physical brain-body-environment, including affective and autonomic aspects of the human body, and the emerging meaning and intentionality of the agent are dependent on the environment and the broader context of interaction [Gal17].

Virtual Fields and Actions Di Paolo describes the concept of the *virtual field*, which contains *virtual actions* [DPBB17]. The functional and structural sensorimotor networks that provide the foundation for cognition consist of active and inactive sub-networks that are dependent on the current configuration of the agent-environment system. The *virtual field* is the "set of concrete dynamic traces and tendencies surrounding a given situation or event as well as neighboring potentialities that have not been actualized" [DPCDJ18]; it describes the dynamic and relational space surrounding an agent within its environment, in which the capacities and tendencies of an agent-environment system exist but are not currently realized. They form a dynamic landscape that surrounds an agent's current trajectory with potential actions and interactions within its environment. Virtual capacities and tendencies are context-dependent and can be potentially infinite (for example, lifting a cup and placing it at various locations in various orientations), but not all of these actions are reasonable or effective for the agent. Possible but not actualized actions in the *virtual field* are *primed* or *inhibited* by neighboring schemata, depending on situated factors [DPBB17]. *Virtual actions* are those actions "that [have] not been fully actualized, and yet have real consequences for the agent and the world" [DPBB17] by influencing what future actions are considered to be beneficial for an agent in a specific situation. This requires the development of a *sensitivity* that is able to "recursively invoke the virtual consequences of actions not yet taken" [DPBB17] to allow for a normative evaluation of possible *virtual*

actions. Some of these possible actions support each other, and some contradict each other, making a careful consideration of the possible outcomes of actions necessary. According to Di Paolo, "all cases of sense-making, and therefore all action and perception, involve constitutively elements of virtuality" [DPBB17].

Enactivism and Human-Computer Interfaces While enactivism has had a large impact on the philosophy and science of cognition, this concept has, so far, only been explored in a limited number of instances in VR and HCI research. Hovhannisyan et al., for example, approach the topic from a theoretical and philosophical perspective, and further illustrate concepts with an artistic IVE [HHS19]. They advocate for an *action-focused approach to VR* that emphasizes the pragmatic dimension of perceptual reality in relation to skill mastery and flow theory. They further distinguish actions based on the physical body (*physical oriented exercises*) and the training of actions which are only possible in IVEs (*virtual oriented exercises*). Rolla et al. describe the term *allusion* to emphasize that experiencing IVEs seems real to the user and should, therefore, not be treated differently from real-world experiences, whereas the fundamental concept of autopoiesis cannot be considered present in IVEs [RVF22]. Cogburn and Silcox argue that IVEs bear the potential to enrich human life by providing new means of interaction as the basis of virtual being [CS14]. Beyond seeing IVEs as an addition to human life, Cantone emphasizes that acting within an IVE may lead to new phenomenological knowledge regarding the sense of *being there*, the *body*, and the *self* [Can22]. Heymaekers emphasizes that, for enactive interfaces, procedural knowledge is an important aspect of interaction that is complementary to other cognitive frameworks, such as symbolic interaction [Ray09]. Often, concepts from enactivism are analyzed and applied on a fundamental level in HCI. For example, Cadoz explores how enactive interfaces serve as a technological counterpoint to cognitive enaction by breaking down interaction into ergotic, semiotic, and epistemic functions and discussing the role of physical feedback loops in human-machine systems [Cad04]. Stoffregen et al. analyze affordances within an environment as essential aspects for designing enactive systems and argue that enactive interfaces should present information that allows users to intuitively perceive available actions through interaction [SBM06]. Kaipainen et al. propose the concept of enactive systems and media, and they argue for a holistic, embodied human-machine coupling based on unconscious physiological responses rather than conscious interface interactions relying on abstract cognitive processes [Kai+11]. It can be summarized that some concepts of enactivism have found their way into VR and HCI research, however, the enactive approach remains a niche topic.

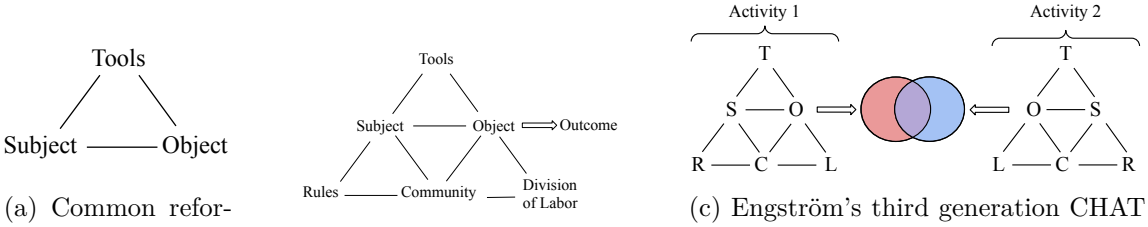
Critique In today's research, enactivism still remains a topic of ongoing debate [WSV17; SW13; DO17], with various strands of thought existing. Radical forms, such as radical enactive cognition [Hut22], reject cognitive representations as fundamental building blocks of all cognition, which makes them, to some degree, incompatible with information processing theory. Instead, the contentless basic mind is proposed as a dominant mode of cognitive operation for basic tasks [Hut22], which exists in both higher-complexity organisms and simple organisms. Other, less radical strands, such as autopoietic enactivism [VTR93], and sensorimotor enactivism [Noë04; DO17] focus on providing a lens on fundamental aspects of cognition and behavior. Stilwell et al. argue that enactivism is crucial to developing a modern understanding of phenomenology and supports qualitative research that explores the complex, dynamic, and context-sensitive nature of sense-making [SH21b]. On the other hand, Meyer proposes that enactivism should be viewed as one approach within the philosophy of nature rather than as a revolutionary perspective on cognitive science [MB22]. Destefano analyzes the debate and proposes that cognition in the enactivist view equals what cognitivists describe as 'behavior' [Des21]. Furthermore, heterogeneous opin-

ions exist on the differences between ecological psychology and enactivism. Read and Szokoleszky [RS20] argue that similarities between these two post-cognitivist approaches exist on a superficial level. Both share the fundamental idea that cognition is the result of dynamic interaction between an organism and the organism's environment. Whereas ecological psychology emphasizes direct perception and action, as well as the lawful constraints in co-dependence between the organism and its environment, enactivism focuses on the individual's sense-making and the emergence of meaning [RS20]. On the other hand, Heras-Escribano [HE21] emphasizes the pragmatic dimension of both approaches and proposes a fruitful combination of both views on interaction as a unified approach to post-cognitivism. Overall, enactivism is still undergoing development, and there is no complete and well-defined enactivism-based research framework for VR and HCI. Still, the proposed implications and fundamental assumptions make enactivism an interesting approach for cognitive science and HCI research.

2.3.4 Activity Theory

The Social Dimension of Action Human behavior is not only influenced by internal factors but also by environmental and social conditions [Ban06]. Human-computer interaction "must be understood in the context of communication and the larger network of equipment and practices in which it is situated" [WF86]. Activity theory, which aims at explaining human action within the social environment, has its roots in early twentieth-century Russian psychology, which differs from the cognitivist approach found in US psychology. The foundation of activity theory consists of two main ideas: i) the social nature of the human mind, and ii) the unity and inseparability of the human mind and activity [CiN16]. These concepts were formulated and elaborated on by various authors, notably, Vygotsky [VC78] and Rubinshtein [Rub46], Engström [Eng01], who popularized the idea of activity theory outside of Russia, and, in the context of HCI, Nardi and Kaptelinin [KN06; Nar96; Nar98; KN97]. In contrast to cognitivist theories, which describe the mind as a complex system that processes information representations in internal processes, activity theory emphasizes that human activity is goal-oriented and influenced by the social and cultural context in which it takes place [Kap96], and human cognitive structures are produced by them [KN06]. It provides a socially acquired perspective as the result of the "enactment of our capacity for attention, intention, memory, learning, reasoning, speech, reflection, and imagination. It is through the exercise of these capacities in everyday activities that we develop; indeed this is the basis of our very existence" [KN06]. A core idea is distinguishing between lower, natural psychological functions, such as memory and perception, and higher psychological functions, which include intention, motivation, and context-dependent initiation [KN06].

Activity Systems As a central tool in activity theory, activity systems are used to describe and analyze human action. In their basic form, they are composed of three elements: the *subject* (S) (the acting individual or group), the *object* (O) (which can also be an immaterial goal), and the mediating *tools* (T) (also referred to as artifacts or instruments). The *subject* is an individual or a group of individuals who carry out the activity, whereas the *object* is related to the goal of the activity. To accomplish the activity, the object and subject are mediated by tools (see Fig. 2.21a), which rely on internal and external components [Kap96]. The conscious use of tools eventually leads to procedures that are then considered internal to the subject, a process called "internalization" [BB03]. Furthermore, the interaction between these elements is influenced by the broader social and cultural context in which the activity occurs. The *community* (C) defines the roles of *subjects* in activities and determines the nature of *objects*. *Subject* and *community* are mediated by *rules* (R), such as norms, values, best practices, laws, and power relations.



(a) Common reformulation of Vygotsky's model.

(b) Engström's general model for activity systems.

(c) Engström's third generation CHAT model for adjunct activities working towards a shared outcome.

Figure 2.21: Three generations of Culture-Historic Activity Theory. Reprint from [Dew+23b], based on [Eng01].

Object and *community* are mediated by *division of labor* (L) (see Fig. 2.21b), which describes the organization of the community necessary to transform an object into an outcome. These components can be visualized as nodes in a network, and the spatial relationship of nodes can be analyzed. Activities have to be seen not in isolation but as coexisting and often working together to achieve a mutual goal (see Fig. 2.21c).

Activity theory, as a tool for analysis, utilizes the idea of *contradictions*. Four different types of contradictions are typically assumed [BB03]: *Primary contradictions* occur within nodes, when values are conflicting. *Secondary contradictions* are observed between nodes, for example, when the community imposes certain rules on how outcomes should be achieved, and a subject cannot comply with these rules in a specific situation. *Tertiary contradictions* describe difficulties in transitions from an established activity system to newer activity systems with more advanced tools, whereas *quaternary contradictions* describe conflicts between neighboring activity systems. Contradictions can further be analyzed at different levels. Cash et al. suggest three hierarchical levels of analysis: the macro-level ("Why do subjects act?"), meso-level ("What do subjects do?") and micro-level ("How do subjects act?") [CHC15], which can be derived from the hierarchical structure of activities with motives, goals, and operations [Kap96]. Another approach is analyzing the sub-triangles of the activity to investigate the mediation between nodes [Mwa01].

Activity Theory in HCI Regarding the use of activity theory in HCI, Clemmensen et al. found five main applications: (1) activity theory itself as object of analysis, (2) activity theory as theoretical influence for subsequent models, (3) activity theory as tool for conceptual analysis and development, (4) activity theory as tool for empirical analysis and (5) activity theory as a framework for design [CiN16]. Applications in the field of HCI are, for example, software development [SR03] and learning in VR [ROS08]. Some expansions have been proposed to account for specific units of analysis, e.g., systemic-structural theory of activity [BK06], which includes predictive models for task performance analysis in activities. Keptelinin lists six interconnected main principles of activity theory that can be described and analyzed in the context of HCI [Kap96]:

1. *The unity of consciousness and activity.* The human mind emerges from interaction with the environment in activities and facilitates the survival of the organism. Therefore, the human mind can only be analyzed in the context in which it is produced.
2. *Object-orientedness.* Socially and culturally determined aspects of the environment in which humans act are equally important as physical and biological aspects and can be treated as objective and meaningful properties.
3. *The Hierarchical Structure of Activities.* Activities can be analyzed on various levels. At the highest level, *activities* are oriented towards *motives* which aim at satisfying

needs. On the next lower level, *actions* are employed to achieve conscious *goals*. On the lowest level, *operations* are performed depending on contextual *conditions*.

4. *Internalization and Externalization*. Internalization describes the transformation of external, inter-subjective actions into intra-subjective mental processes through social interaction. Conversely, internal mental processes lead to external actions.
5. *Mediation*. Tools mediate human activity. The internalization of cultural tools shapes the mental development and related actions. This encompasses external, physical tools as well as internal tools, such as concepts and symbols.
6. *Development*. Activities are constantly changing. To fully understand activities, it is necessary to understand their context and historical development.

PART II

SUPER-NATURAL INTERACTION

CHAPTER 3

MAGIC AND TECHNOLOGY

3.1 Computer Magic

Perceiving Technology as Magic In 1857, *The Magician's Own Book* was published, which contained magic tricks and experiments based on novel scientific advancements of that time, such as electricity, magnetism, and chemistry. Scientists were "yet ignorant of [their] nature" [Arn+57] and, therefore, these novel effects were used to produce 'magic' for entertainment purposes. More than a century later, in 1973, Freedman published a paper titled: "Computer Magic" [Fre73]. He reported a form of magical thinking in computer science students who had to create program code on punch cards. Examples of their magical thinking were the idea of 'mana,' an unexplainable power contained in specific cards that were able to solve a task, 'rituals,' the conservation of the order of steps of procedures required to obtain a result, and 'name magic,' the naming of variables in specific ways that support the program code. The belief in the effects of performing this computer magic and combining different magical ideas into a set of magical best practices influenced how humans interacted with computers [Fre73]. Today, in contemporary engineering, the term 'magic' is still used to describe the properties of interfaces in HCI. For example, during the presentation of the iPhone in 2007, the novel multi-touch technology[8] of the first-generation Apple smartphones was described as "... phenomenal. It works like magic" [Inc13].

Intentional Magic Design Purposefully implementing scientifically unexplainable and non-realistic behavior can also be viewed as an approach to designing systems. In 1986, Smith presented the Alternative Reality Toolkit (ART) [Smi86] to research the tension between what he called *literal* and *magic* interaction techniques in desktop applications. ART allowed, for example, the implementation of behavior that contradicted reality, such as inverting the direction of gravity. Smith defined magical features "to be those capabilities that deliberately violate the metaphor in order to provide enhanced functionality" [Smi86]. A different notion of 'magic' was presented in the Magic User Interface toolkit for AMIGA systems in 1993 [Stu98]. Here, magic referred to the ease and simplicity of creating custom WIMP environments that were controlled in conventional ways. Both dimensions can be found in Jacob's general framework for *Reality-Based Interaction*, in which he describes the tradeoff between features of reality (*Naïve Physics, Body Awareness & Skills, Environment Awareness & Skills* and *Social Awareness & Skills*) and beneficial features of interaction (*expressive power, efficiency, plasticity, ergonomics, accessibility* and *practicality*) for post-WIMP interfaces [Jac+08a]. In this framework, realism can be intentionally reduced to enhance power or to increase the simplicity and accessibility of interaction.

Magic in IVEs IVEs are an interesting area of research for magic interaction as they enable users to perform actions that are not constrained by the real world, while other aspects of reality can be faithfully simulated. The visionary essay *The Ultimate Display*

[8] (Multi-)touch screen technology had been available for several decades (see [Joh65] and [LBS85]). However, the iPhone can be seen as the development that paved the way for the widespread use of this technology in large parts of society.

[Sut65] from 1965, which can be considered highly influential in VR interaction research, already presented the idea of not "follow[ing] the ordinary rules of physical reality with which we are familiar" [Sut65] as a possible application and an intriguing concept for IVEs. Walser described in 1990 so-called *spacemakers*, the creative designers of virtual environments, who possess the *supernatural powers* to override *natural laws* that contrast realistic, *terrestrial* interaction [Wal90]. A similar idea was articulated by Stephen Ellis in 1991, who described how the arbitrary nature of virtual environments enables supernatural interaction in contrast to natural interaction, which uses the real world as a frame of reference [Ell91]. Jaron Lanier describes *absolute physics* in VR and MR systems as the implementation of rules that only apply within such a system, which can be utilized for new forms of interaction [Lan88].

Various authors have further elaborated this concept. For example, Shneiderman proposes that, by not recreating reality, 3D environments allow for "superhuman capabilities such as faster-than-light teleportation, flying through objects, and x-ray vision" [Shn03] to enable a simpler and more productive or compelling interaction, sometimes even including non-realistic, "super-natural tools such as magic wands" [Shn03] that provide these enhanced functionalities to users. Appropriate content and compelling ways of interaction in such an enhanced environment have the "potential for novel social, scientific, and commercial applications if designers go beyond the goal of mimicking 3D reality" [Shn03]. Kulik emphasizes that instead of recreating reality, it is more important to provide an interaction that is "effective, fun, or both in the best case" [Kul09]. Interaction can draw inspiration from "popular science-fiction stories, common knowledge about technological achievements, and the arbitrary combination of aspects from previously learned patterns" [Kul09] to provide comprehensible metaphors to users. Kulik further presents imagination-based concepts that contradict the experience of the real world, such as *suspension of naïve physics, geometric scaling, motion scaling, automation, magic spells*, and *mode changes*. [Kul09]. Bowman contrasts two design approaches for 3D user interfaces: i) natural interaction, in which the interaction from the real world is faithfully recreated in IVEs, and ii) magic interaction, which he characterizes as "intentionally less natural, or [magic techniques that] might enhance natural interactions to make them more powerful" [BMR12]. He emphasizes that magic interaction "can make performing tasks in the virtual world easier than in the real world" [BMR12], especially when designers follow a so-called *hyper-natural* approach that implements familiar and natural feeling interaction, which does, however, not fully replicate real-world interaction to provide "users with enhanced abilities that improve performance and usability" [BMR12]. Similarly, Serpi et al. describe *hyper-natural* interaction techniques as "natural-feeling actions to control "magical" actions, whose goal is making the manipulation simpler going beyond real-world constraints" [Ser+18]. LaViola et al. [LJ+17] describe magic as a possible design strategy for 3D interfaces, which enables a user to overcome human limitations by enhancing cognitive, perceptual, physical abilities, and motor control, which ultimately leads to a "better reality" compared to just simulating or adapting the real world in IVEs [LJ+17].

Magic interaction techniques can be applied to various types of interaction, both cognitive and physical. Abtahi et al. focus on sensorimotor transformations of movements for interaction in IVEs and distinguish between *reality-based* (complete symmetry to reality), *illusory* (interactions that are different from reality but below a perceptual threshold), and *beyond-real* interaction (interactions that are perceived as not possible in the real world) [Abt+22]. Willet et al. apply Kirsh's concept of *epistemic* and *pragmatic actions* [KM94] to magic interaction techniques and present examples of comic book heroes as inspiration for *epistemic superpowers* as well as *pragmatic superpowers* [Wil+21]. In their framework, *pragmatic superpowers* enable users "to actively manipulate things, people, or phenomena in the world" [Wil+21], whereas *epistemic superpowers* allow them "to gain knowledge

about the world without necessarily changing it" [Wil+21]. For the latter, they present *dimensions of empowerment* and envision future directions of *empowering visualizations*. Sadeghian and Hassenzahl describe the concept of "superpowers" in designing interactive applications [SH21a]. They distinguish between *perception*, *cognition*, and *action* and provide the *SuperPowerLab* toolbox for Unity3D as a tool for designing IVEs.

Frameworks Containing Magic These conceptual ideas regarding magic interaction have been incorporated into several frameworks that relate magic and real-world interaction. Slater and Usoh describe the continuum between *mundane* and *magical* to distinguish between faithfully reproducing reality and interaction that cannot occur in reality, such as flying and teleportation [SU94]. Based on this magical-mundane continuum, Nielsson et al. present a taxonomy for Walking-In-Place locomotion techniques that includes a *plausibility* dimension [NSN16]. In the VRID framework, Tanriverdi and Jacob distinguish between *physical behavior* and *magical behavior* as properties of virtual objects in IVEs [TJ01]. *Magical behavior* contains all actions, behavior, and states "which are rarely seen, or not seen at all in the real world" [TJ01]. By decomposing complex behavior into magical and physical behaviors, the VRID framework aims to improve communication between designers and developers, as well as simplify development by utilizing reusable components. Gladden systematically analyzes different types of "magical practice" employed by users in HCI, such as the "use of magic words" or the "use of specially prepared ritual implements" [Gla19]. He emphasizes that "magicality is a complex phenomenon that has the potential to both improve and damage usability and user experience in diverse ways" [Gla19].

Synonyms When reviewing the literature on the topic of magic interaction in HCI, the missing delineation of precise terms to describe these interactions is quickly observed. Researchers seem to often rely on personal preference when choosing specific terms and rarely state the reasons for naming an interface in a specific way. This can, ultimately, blur the meaning of terms, leading to confusion or necessitating clarification of what an author actually means. To illustrate the diversity in this usage of terms, the results of full-text database queries using the ACM Digital Library Full-Text Collection[9] and IEEE Xplore[10], which are important digital libraries for research in HCI and VR, are summarized in Table 3.1, which shows the raw number of publications that use a corresponding term[11] in combination with 'virtual reality'. The search terms employed in the queries originate from an informal literature retrieval and, presumably, do not represent a comprehensive list of all possible terms that are used in research and design. Furthermore, the query results were not further analyzed, and, therefore, not all publications necessarily include some form of magic interaction, as these terms can also be utilized in different contexts.

For every magic-indicating term, at least one publication can be found (see column 'Example' in Table 3.1), which, on a superficial level, describes a similar design approach to VR interaction. Based on this finding, it follows that potentially all of these terms can be used to search for related literature. However, without further analysis, it is difficult to estimate the achievable precision and recall corresponding to each term. Considering the amount of published research, it is apparent that the use of general terms for literature retrieval, such as, 'augment*,' 'hyper*,' 'empower*,' and 'enhanc*,' yields a quantity of literature that cannot easily be handled, and which presumably not necessarily achieves an adequate precision for identifying 'magic' interaction techniques. On the other hand, more specific terms, such as 'super-natural,' 'superhuman,' 'superpower,' and 'beyond-real,' yield only

[9] https://dl.acm.org/

[10] https://ieeexplore.ieee.org/

[11] An '*' is used to represent an arbitrary number of other letters, for example, to account for 'magic', 'magical' and 'magician' using only the single expression 'magic*'.

Term	ACM	IEEE	Example
Beyond-Real	25	88	"interactions ... that go beyond our experience of reality, which I have called "beyond-real interactions"."abtahi2019m
Superpower*	80	82	"Using Superpowers in Virtual Reality to Encourage Prosocial Behavior" [RBB13, publication title]
Super*natural	89	81	"... to perform the "supernatural" task of growing your arm in VR." [EWK18]
Superhuman*	104	105	"... using technology to surpass the physical and cognitive restrictions of our bodies and enabling superhuman senses and abilities." [Kun+17]
Exaggerat*	777	1012	"... to test if exaggerated movement brings benefits in a martial arts kicking task ... " [Gra+18]
Empower*	1.542	3.288	"... tools can give the people who use them a sense of objective or subjective empowerment ..." [Wil+21]
Magic*	2.756	2.259	"Capabilities that violate the metaphor in order to provide enhanced functionality might be called magical." [Smi86]
Hyper*	4.838	9.415	"... enhancing users' real-world abilities with a hyper-natural interaction technique." [NB15]
Augment*	12.881	26.892	"... augment the physical, cognitive, and perceptual capabilities of users" [SH21a]
Enhanc*	11.641	35.630	"... providing users with enhanced abilities that improve performance and usability." [BMR12]

Table 3.1: Raw number of research query results using magic-related terms in the context of VR for ACM DL and IEEE XPlore (As of Aug. 9th 2023).

a small fraction of the overall amount, which implies a low recall capability. Some terms yield a medium amount of literature, such as 'magic*' and 'exaggerat*', but in each of these cases, the specific use of the term cannot be inferred without a full-text analysis. In all three cases, it is not possible to be confident that the retrieval of all relevant research work was successful using only a single term or a combination of several terms. Furthermore, it is not possible to rule out that further terms exist that were not considered in this screening step.

3.2 Literature Review on 'Super-Natural'

In this thesis, the focus lies on the term 'super-natural' interaction as a conceptual foundation. As presented in the previous section, 'super-natural' is not an exclusive term to describe magic interaction techniques, and delineating this term from other terms is not possible without further analysis. Furthermore, it is unclear what meaning is conveyed by authors when they use this term, and if all uses of this term imply the same meaning. The central question of this review is research question **R1**:

R1: How is super-natural interaction described in research literature?

3.2.1 Methods

To answer **R1**, a literature review was performed in which text passages of research articles on 'super-natural' in VR and related technologies interaction were collected and analyzed regarding the individual use and meaning. The methodology employed in this literature review combines both semi-automated text extraction techniques and manual assessment to provide a comprehensive analysis and synthesis of existing research.

In the initial phase of this review, a semi-automated text extraction was performed to gather relevant text passages from the sources in the literature corpus. First, the retrieved articles were converted from PDF format to plain text files using the Python package PyPDF2. The plain text files were manually assessed to check for errors in the conversion. In case of failed conversions, Poppler[12] was used to convert the PDF files to PNG image files (with a resolution of 600 dpi), and Tesseract-OCR[13] was employed to convert PNG files into one text file for each article. After conversion, the Python packages *ntlk* and *sklearn* were used to perform a concordance analysis to retrieve the context (600 words) of occurrences of the terms 'supernatural', 'super natural,' and 'super-natural' for each article in the literature corpus to enable the identification of key themes, terms, and relevant text passages. Following the semi-automated text extraction and concordance analysis, each article's PDF file was manually assessed to refine and contextualize the findings and to produce descriptive statistics for the literature corpus.

3.2.2 Literature Retrieval

Systematic Database Queries To retrieve a sufficient literature corpus, a search was conducted in the ACM Digital Library[14] and in IEEE Xplore[15]. The search was conducted as a full-text search to retrieve publications that contain variations of the term 'super-natural' ('supernatural' and 'super natural') in combination with 'virtual reality' (and the related terms 'vr', 'virtual environment*', 'augmented reality,' 'AR,' 'XR,' 'mixed reality'). The search was limited to journal and conference papers of all types, e.g., extended abstracts, work-in-progress articles, short papers, and full papers. The queries were carried out on February 26th, 2023, covering all publications until December 31st, 2022, to capture an up-to-date impression of the utilization of this term in research. For transparency and clarity, the literature retrieval incorporates essential aspects of the PRISMA guidelines [Pag+21] and includes the PRISMA flow diagram (see Figure 3.1).

The resulting **IEEE Xplore** query was:

("Full Text & Metadata":"Supernatural" OR "Full Text & Metadata":"super natural" OR "Full Text & Metadata":"super-natural") AND ("Full Text & Metadata":"VR" OR "Full Text & Metadata":"virtual reality" OR "Full Text & Metadata":"virtual environment" OR & Metadata":"virtual environments" OR "Full Text & Metadata":"XR" OR "Full Text & Metadata":"augmented reality" OR "Full Text & Metadata":"AR" OR "Full Text & Metadata":"mixed reality")

The corresponding query for **ACM Digital Library** was:

AllField:("Supernatural" OR "super natural" OR "super-natural") AND AllField:("VR" OR "virtual reality" OR "virtual environment" OR "virtual environments" OR "XR" OR "augmented reality" OR "AR" OR "mixed reality")

[12] https://github.com/freedesktop/poppler
[13] https://github.com/tesseract-ocr/tesseract
[14] https://dl.acm.org/
[15] https://ieeexplore.ieee.org/

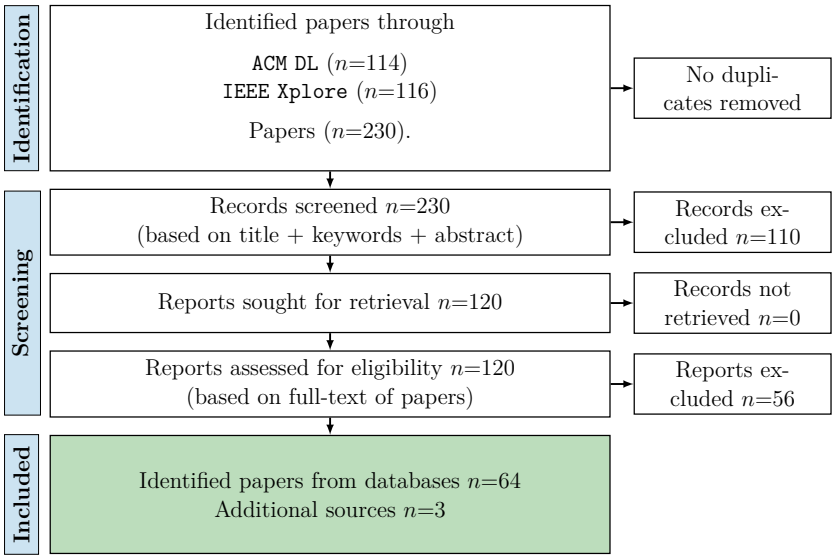


Figure 3.1: PRISMA flow diagram [Pag+21] describing the literature retrieval.

Eligibility The queries yielded 230 publications (ACM: 114; IEEE: 116), which were further analyzed to determine eligibility. First, in a screening phase, each publication's title, keywords, and abstract were analyzed. Three criteria had to be met for inclusion: The research article investigates (1) interaction or user-specific aspects (2) in the context of IVEs or related technologies, and is (3) written in English or German. From the initial corpus of 230 papers, 120 remained for further analysis (ACM: 68; IEEE: 52) after the screening phase. All of these were retrievable. The semi-automated extracted passages were assessed to decide on inclusion with the following additional criteria: (4) 'super-natural' is used as a descriptive term for this specific research work, and (5) no matching article has already been added to the literature corpus[16]. If it was not possible to decide on inclusion based solely on the extracted text passages, the full text was reviewed. After the eligibility assessment, 64 papers (ACM: 39; IEEE: 25) from the database queries remained in the final literature corpus.

Additional Literature Sources Additionally, the Internet was informally searched for publications containing explicit descriptions of 'super-natural' HCI in IVEs. One publication is a PhD thesis by Nguyen [Ngu14], which describes 'super-natural' object interaction and locomotion in IVEs. The second source is a PhD thesis by Nabiyouni [Nab15], and the accompanying paper by Nabiyouni and Bowman [NB15], treated as one additional source for analysis. Both talks by Steinicke ([Ste17b] and [Ste16b]) correspond in content and position regarding 'super-natural' interaction to [Ste17a], which was already obtained from the database queries. Again, these three sources are treated as one combined source in the analysis. Last, the PhD dissertation by Lubos [Lub18], which analyzes ergonomic interaction and usability factors of user interfaces was added. Table 3.2 displays the additional sources that were identified and incorporated into the literature corpus. With these three additional publications, the resulting literature corpus contains 67 publications (see appendix B for a complete list).

[16] In some cases, a similar or updated version of a paper has been submitted with a different title. In these cases, the publication containing more details was selected, and the less detailed article was omitted.

Author	Year	Reference	Type	Title
Nguyen	2014	[Ngu14]	PhD thesis	Proposition of new metaphors and techniques for 3D interaction and navigation preserving immersion and facilitating collaboration between distant users
Nabiyouni and Bowman	2015	[NB15]	Paper	An Evaluation of the Effects of Hyper-Natural Components of Interaction Fidelity on Locomotion Performance in Virtual Reality
Nabiyouni	2015	[Nab15]	PhD thesis	An Evaluation of the Effects of Hyper-Natural Components of Interaction Fidelity on Locomotion Performance in Virtual Reality
Steinicke	2016	[Ste16b]	Talk	Super-Natural User Interfaces for the Ultimate Display.
Steinicke	2017	[Ste17b]	Keynote	Fooling your Senses: (Super-)Natural User Interfaces for the Ultimate Display.
Lubos	2018	[Lub18]	PhD thesis	Supernatural and Comfortable User Interfaces for Basic 3D Interaction Tasks

Table 3.2: List of publications added to the literature corpus that were retrieved using an informal Internet search.

3.2.3 Analysis

Use of Super-Natural The first source in the investigated literature corpus that used the term 'super-natural' was published in 1999. Until 2014, the term was used only in rare cases, and for many years, not a single paper was published that contained this term. The use frequency increased after 2013, however, the term can still be considered rarely used in contemporary research (see Fig. 3.2).

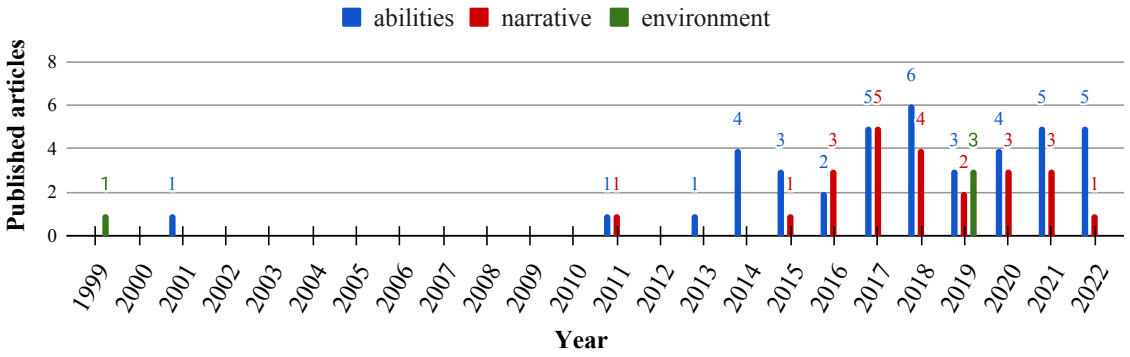


Figure 3.2: Number of publications per year for super-natural abilities (blue), super-natural narratives (red), and super-natural environments (green).

The most common way of spelling the term is 'supernatural' (61 articles), whereas 'super-natural' (5 articles) and 'super natural' (1 article) are relatively uncommon in compari-

son[17]. Only one publication in the literature corpus utilized 'supernatural' as a keyword to improve the retrieval of the research [Mar+18], two publications used 'magic' [YLO17] or 'magical' [SBTP20], and two others 'superhuman' ([And+18; Buc+18]). 'Supernatural' appeared in the title of four publications ([EWK18; Kru+16; Mar+18; Bec+19]. In summary, indicating the super-natural properties of a presented interaction technique is not common practice.

Super-Natural Abilities / Narratives / Environments Three concepts can be distinguished in the literature corpus regarding the user of 'super-natural'. The first and most dominant interpretation (40 publications) of 'super-natural is the implementation of new means of interaction that users utilize to solve specific tasks. In this thesis, these are referred to as *super-natural abilities*. The second main interpretation (23 publications) involves the utilization of narrative elements with unexplainable characteristics, which is referred to as *super-natural narratives* in this thesis. A third interpretation is the construction super-natural environments (4 publications) that provide properties that differ from reality.

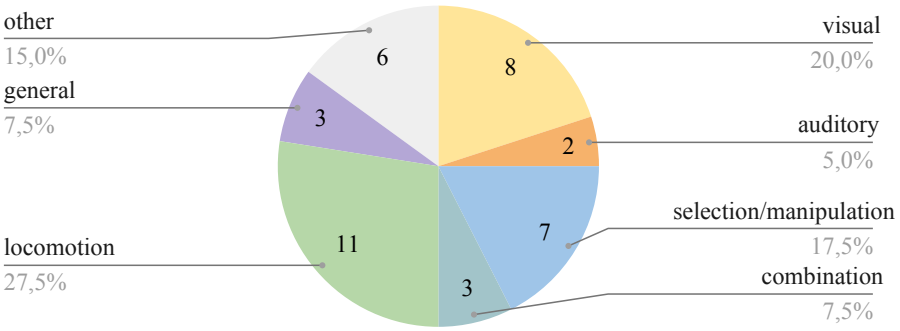


Figure 3.3: Main topics of research regarding super-natural abilities in the literature corpus.

Super-natural abilities enhance senses or physical actions, or they change the user's virtual body (see Table 3.3). The largest group of abilities contains locomotion techniques [PMW22], especially forms of teleportation [Fis+17; Pit17; HL19; SSH20; RZ21; HMS22], but also flight [Kru+16] and other real-world inspired techniques [Nab15] are investigated. Object selection and manipulation are also of high interest, especially as some form of telekinesis [LBS14; Hay+15; Pfe+17; Spe+18b; EWK18; Woź+21; CS21; TS21]. In some cases, a combination of various techniques is considered in a single research work [Gau+14; Ngu14; Lub18]. In this literature corpus, eight publications propose changes to vision as super-natural interaction techniques, for example, to reduce occlusions (e.g., 'x-ray vision') [Wan+19; WP17; CS01; GBB20] or to enable users to see more details in an environment [LDH13; Woź+21]. Another strategy is to add visual elements instead of increasing visual abilities. Laera et al. propose the use of additional super-natural visual elements to display important information to users in an AR sailing navigation application [Lae+20]. Kush proposes projecting digital information into the real environment to enable beneficial applications that would have been "considered to be supernatural or tantalizing imagination" [Kus15] a few years ago. Poretski and Tang [PT22] propose such overlays for games and also AR applications. Similar to visual enhancement, auditory perception can be enhanced by improving the auditory capabilities [SHP18], or by providing new ways of interacting with sound [NS19]. An implicit topic that overlaps with super-natural abilities is the implementation of super-natural modalities, for example, brain-computer interfaces [Woź+21] or

[17] In this thesis, the spelling 'super-natural' is used to express specific properties discussed in chapters 4, 5, and 6.

Topic	Count	References
Locomotion	11	[Nab15; Kru+16; Fis+17; Pit17; HL19; Zha+20; SSH20; RZ21; PMW22; HMS22; Hay+15]
Selection & manipulation	7	[Pfe+17; Spe+18b; EWK18; Zha+20; Woź+21; CS21; TS21]
Combination	3	[Gau+14; Ngu14; Lub18]
Visual	8	[CS01; LDH13; Kus15; WP17; Wan+19; Lae+20; GBB20; PT22]
Auditive	2	[SHP18; NS19]
General	3	[MSS14; Ste17a; Sim+22]
Other	6	[Apo+12; Gug+16; Spe+18a; Buc+18; Kre+21; Yu+22]

Table 3.3: References and count of publications for each type of investigated super-natural abilities.

haptic sensations [Rak+20; Mar+18]. Super-natural abilities also enable actions that are not possible at all in reality. Examples in the investigated literature corpus include the creation of objects such as walls [Buc+18] and sounds in mid-air [NS19]. Users can also change their body, for example, by increasing or decreasing the size [Zha+20; Kre+21], or controlling additional body parts [LBS14]. Other types of interaction are text-input [Spe+18a], communication [Apo+12], 3D annotation [Yu+22], and input-technology [Gug+16]. Some of the included research articles also present general ideas and theories on implementing and conceptualizing super-natural interaction [Ste17a; MSS14; Sim+22].

Super-natural environments, on the other hand, are described to contain "supernatural properties which expand or replace the laws of physics ..." [Bec+19], which makes them "detached from reality" [Cos+19]. In some cases, these super-natural properties allow for super-natural interactions [Cos+19; Med+19]. Examples of super-natural environments and components of environments are multiscale IVEs [Med+19], virtually connected IVEs (SEAMS) [SS99], and environments that change the effect of gravity [Bec+19; Cos+19].

Finally, super-natural narratives encompass the interaction with entities, either objects or other simulated characters in the IVE that are perceived as super-natural. In some cases, these are paranormal forces used to provoke reactions in users, for example, haptic sensations of touch [Isr+15; Mar+18; Rak+20; Pit+21] or temperature [LLK20]. In other cases, the experience can be described as super-natural, for example, autonomously opening cabinet doors in response to the user [Hva+17a] and other unexplainable events [SBTP20; Byr+22]. Sometimes, the setting and theme of an IVE induce a super-natural feeling [VS17; Hva+17b]. Furthermore, the interaction with super-natural entities such as ghosts [Fur+19] and other fantastic beings that do not exist in reality [BMM16; Min+21] is also described as super-natural, or objects possessing super-natural qualities (e.g., Ouija boards [BB19] and magical wands [TI+21]). Interaction can be aligned with fictional content [Mit+17] as well as cultural traditions and popular beliefs [BX12; KAN16; YLO17; And+18; Rag+20; BB19]. For super-natural narratives, it is important to provide consistent rules even when rules of the real world are discarded to facilitate the understanding of the intended interaction in users [BX12; MS+18].

A taxonomy for super-natural interaction can be derived from the applications presented in the literature corpus (see Fig. 3.4). Considering the small amount of literature in this review, it is probable that this taxonomy is not comprehensive regarding all possible ways of making interaction 'super-natural.' However, it still provides an overview of various super-natural strategies.

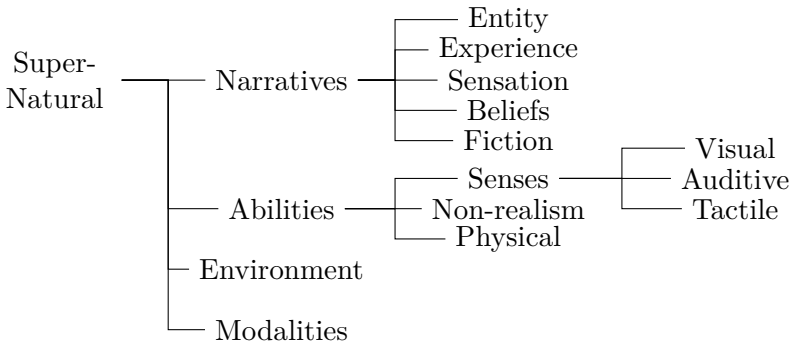


Figure 3.4: A taxonomy derived from identified forms of 'super-natural' in the literature corpus.

Thematic Analysis of Super-Natural Abilities In the subsequent analysis, super-natural abilities are further investigated as the primary topic of this thesis. The initial point for analyzing the properties of super-natural abilities were the framework by Nabiyouni and Bowman [NB15; Nab15], and the definitions provided by Steinicke [Ste17a; Ste16b; Ste17b] and Lubos [Lub18]. Both approaches share similarities, but they also differ in certain aspects. First, the central statements by Nabiyouni and Bowman are presented, which describe super-natural interaction techniques in the context of locomotion interfaces for IVEs:

> "[D]esigners can make the choice to reject the real-world metaphor altogether and design a non-natural interaction technique with low levels of interaction fidelity. This could take the form of a non-natural technique in which the designer simply determines an efficient mapping between the input and desired actions, such as the joystick or keyboard controls used in many video games. On the other hand, designers can create super-natural techniques that go far beyond reality to provide users with unrealistic superpowers." [NB15]

> "With non-natural techniques designers intend to provide virtual actions similar to the real world (e.g., walking) while the interaction method is far from the corresponding real-world action [...]. On the other hand, designers can create super-natural techniques that go far beyond reality to provide users with unrealistic superpowers. In super-natural techniques, the designer typically puts a "story" around the technique and provides the user with abilities beyond real-world actions [...] while the corresponding physical action is not natural. In both the non-natural and super-natural approaches, developers have tremendous freedom to design effective techniques without the constraints of the real world." [Nab15, p. 5]

In their descriptions, Nabiyouni and Bowman explicitly contrast 'super-natural' interaction with both real-world interaction and non-natural interaction techniques and emphasize the use of *unrealistic superpowers*. They also include the utilization of narrative elements in their description to distinguish between non-natural interaction techniques, which only focus on providing efficient mapping, and super-natural interaction techniques. Using

McMahan's Framework for Interaction Fidelity Analysis (FIFA) [MLP16], they objectively analyze the degree of faithfully replicating real-world actions and further distinguish between 'super-natural' and 'hyper-natural' techniques as two distinct levels of interactions with low fidelity. *Hyper-natural* techniques are described in their framework as techniques that are based on natural interactions, such as walking in an environment or grabbing an object, which are enhanced by reducing fidelity. *Super-natural* techniques, on the other hand, are completely detached from real-world interaction (which produces a very low fidelity in the FIFA framework) and provide entirely new ways of interaction.

In the approach by Steinicke, first, natural user interfaces are described as interfaces that can be operated by users who only understand *intuitive physics*, which is described as "basic understanding of very simple physical laws" [Ste16b]. Based on this, he defines super-natural user interfaces (SNUI)s as:

> "**SNUI** is an **interface** that can be operated by a user who only knows about **intuitive physics**, but which is not limited to **physical reality**." [Ste16b]

Lubos references Steinicke's definition in his dissertation. Expanding upon the concept of characterizing interfaces as 'natural' when they are based on what he describes as "natural human mechanisms", such as, "grabbing, walking, touching, speaking, looking" [Lub18], he describes super-natural interfaces as:

> "Supernatural user interfaces (SNUIs) are interfaces which are still inspired by the ways humans interact with one another or with their environment, but not limited by it. SNUIs permit actions which are not possible in the physical world. Examples would include teleportation or floating interface elements. Since virtual realities allow developers or users to set the rules within a world, the way users interact with virtual environments (VEs) can also be supernatural. Natural interaction can still inspire these interactions, however, they are less limited by natural constraints." [Lub18, p. 3]

He further proposes a definition that slightly differs from Steinicke's:

> "A user interface is called **supernatural**, when it is based on natural human mechanisms, but not limited to laws of physics or real-world constraints." [Lub18, p. 13]

The definitions by Steinicke and Lubos share with the framework by Nabiyouni and Bowman the idea of not limiting interaction to the constraints of reality or limitations of the human body. However, they neither include what Nabiyouni and Bowman call "unrealistic superpowers" [NB15; Nab15] nor other forms of empowerment, nor story elements in their definition. Instead, they focus on 'natural' interaction as an important component, in the case of Steinicke, *intuitive physics* [Ste16b], in the case of Lubos, *natural human mechanisms* [Lub18], which is not emphasized in the framework by Nabiyouni and Bowman.

These central ideas about super-natural interaction can be identified in diverse combinations in the literature corpus. For example, Mostafa et al. describe "the superhumans metaphor with the aim of empowering designers and users to think about VEs as unique experiences providing supernatural abilities and elastic interactions" [MSS14]. In their view, these abilities are "well beyond [the users'] real-world experience" [MSS14] and can include narrative elements. Gugenheimer et al. contrast natural interaction concepts and ""super natural" interaction where users can interact and manipulate the virtual environment with little physical effort and enable interactions beyond human capability" [Gug+16]. They furthermore report that participants in a study of their proposed *FaceTouch* interface stated that the interaction was fast to learn and also intuitive and natural to use.

Apostolopoulos et al. describe that for communication in IVEs "in addition to natural ("as good as being there") communication enabled by immersion, supernatural ("better than being there") communication will also be important. For example, it will be possible to look or listen across large distances, rewind or accelerate time, speed through space, record experiences, or provide translation abilities" [Apo+12]. For Yu et al., providing the non-realistic ability to scale 3D objects to draw annotations with higher precision leads to a "supernatural precise interaction" [Yu+22]. Speicher et al. write that *super-natural metaphors* can be employed for object selection and manipulation to "overcome limitations in the tracking space or anatomical constraints" [Spe+18b]. However, Nguyen describes that "in the design of manipulation techniques in immersive virtual environments, a trade-off has to be made between the naturalness and the efficiency of the interaction technique" [Ngu14, p. 66]. Choosing a super-natural tool may lead to better efficiency [Ngu14, p. 66], but in some cases, a natural interaction "can offer greater performance and usability and stronger feel of presence" [Ngu14, p.32].

Overall, teleportation in IVEs was referenced multiple times as an example of super-natural interaction [Pit17; Fis+17; SSH20; RZ21; HMS22]. Pittarello describes that "teleporting doesn't belong to the normal human experience" [Pit17], but adequate means of input are capable of conveying "the sense of supernatural powers, yet maintaining simplicity and avoiding the need of external devices" [Pit17]. Fisher et al. [Fis+17] describe the category of "super-natural powers; a magical approach, such as teleportation" [Fis+17] for locomotion as a possible low-fidelity strategy for locomotion besides "non-natural tactics" [Fis+17]. Sayyad et al. [SSH20] propose that teleportation can be especially beneficial as a locomotion technique in large environments. In a similar way, Han et al. [HMS22] emphasize the use of teleportation in VR systems with limited physical space. Sayyad et al. state that in some cases, the "... improved speed and flexibility of a 'super-natural' technique such as teleportation, could yield some benefits ..." [SSH20] even though, in their study, natural walking was preferred by most participants. Teleportation as an *artificial* locomotion technique was considered by some of their participants to be easy to use but performed worse than natural walking in terms of building a mental map, overall usability, and simulator sickness [SSH20].

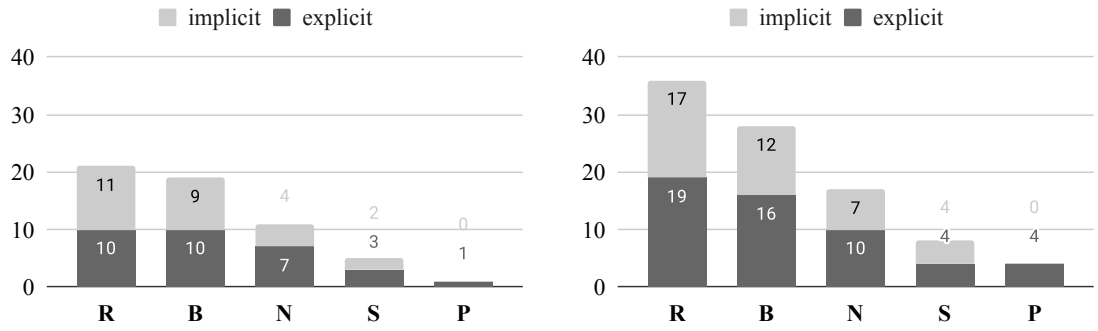
For some researchers, both immersion and presence are interesting aspects of super-natural interaction. Wang et al. propose that *super-natural powers* "should be designed such that the user is convinced of the immersion in the virtual world, which they explore with the benefit of a superpower" [Wan+19]. Wu et al. even propose that developers should not "attempt to hide the added supernatural visualization capability" [WP17] and enable users in their prototype system to change their super-natural visualization (looking through walls) "intuitively, with minimal or no interface manipulation" [WP17]. Rieke and Zielasko describe teleportation as a "convenient, easy, effortless [and] practical" [RZ21] locomotion technique that "does not really have any real-world equivalent (besides science-fiction depictions of teleporters)" [RZ21] and may interfere with presence and immersion.

Two sources in the literature corpus provide no further information on the meaning of the term 'super-natural'. In their systematic literature review of different approaches for locomotion techniques, Prinz et al. [PMW22] derive a taxonomy for locomotion techniques from Nilsson's PhD thesis [Nil15] that contrasts natural walking as a high-fidelity way of locomotion to low-fidelity locomotion, including super-natural, magical, and non-natural techniques, without further discussing differences between these terms. Krupke et al. present a "supernatural flight simulator" [Kru+16], without providing details on the properties that make this prototype 'super-natural' in this paper.

Overall, five central themes of 'super-natural' interaction have been identified:

(R) Removing constraints of the real world or human body.
(B) Allowing 'better' ways of interaction ('superpowers').
(N) Based on 'natural' ways of interaction. Intuitive and easy to use.
(S) Inclusion of narrative and story elements.
(P) Reduced immersion and presence.

None of the reviewed articles references the definitions or framework to express what 'super-natural' means in each individual case. Instead, authors mostly rely on short descriptions or examples to clarify what 'super-natural' is intended to characterize. In many cases, no explicit statements are included, and the meaning has to be inferred from descriptions of the interaction. However, it is possible to analyze the publications regarding the frequency of using implicit (e.g., participants reporting the ease of use as a result of a user experiment) and explicit statements (e.g., stating the super-natural interaction techniques enable non-realistic experiences as a property of the interaction technique) that include themes in their concept (see Fig. 3.5).



(a) Number of publications featuring a specific theme from 1999 to 2018 ($n = 23$).

(b) Number of publications featuring a specific theme from 1999 to 2022 ($n = 40$).

Figure 3.5: Frequency of included themes (beyond-real **R**, better **B**, natural **N**, story elements **S**, and effects on presence and immersion **P**) in the literature corpus.

For this literature corpus, enabling users to perform actions that are not possible in the real world (**R**) is the most described property of super-natural interaction (90%). Many publications also describe enhancing the capabilities of users (**B**, 70%) and providing natural and simple ways of interaction (**N**, 42.5%). Story elements (**S**, 20%) and effects on immersion and presence (**P**, 10%) are reported less frequently. A 2x5 chi-squared test revealed that the distribution of statements (both implicit and explicit) regarding specific themes in the reviewed literature corpus does not differ significantly ($X^2(4, N = 92) = 3.385, p = .496$) between publications published until 2019 and publications published from 2019 to 2023.

Furthermore, the combination of themes provides valuable insight. The conjunction of explicit and implicit statements (see Fig. **??**) shows that the most commonly observed combinations are the groups 'RB' (non-realism and enhanced functionality) and 'RBN' (non-realism, enhanced functionality, and simple use), followed by 'R' (non-realism). Two publications did not further discuss which properties make interaction 'super-natural' (group '0'). However, if only explicit statements are considered, the distribution changes (see Fig. 3.6). In 10 publications, the authors did not explicitly describe what they mean by using the term 'super-natural,' which implies a use in the form of a description rather than attributing the interaction to a specific class of interaction techniques. Often, only

one component is explicitly discussed (e.g., non-realism), whereas the influence of other properties is not directly mentioned. In both cases, explicit and implicit statements, and explicit statements only, it should be noted that not mentioning a specific property does not mean a negation. For example, in publications assigned to the group 'RB,' no statement was identified regarding the ease of use ('N'), so interaction techniques that are easy to learn, as well as those that are hard to learn, could be considered implicitly included in the group 'RB.'

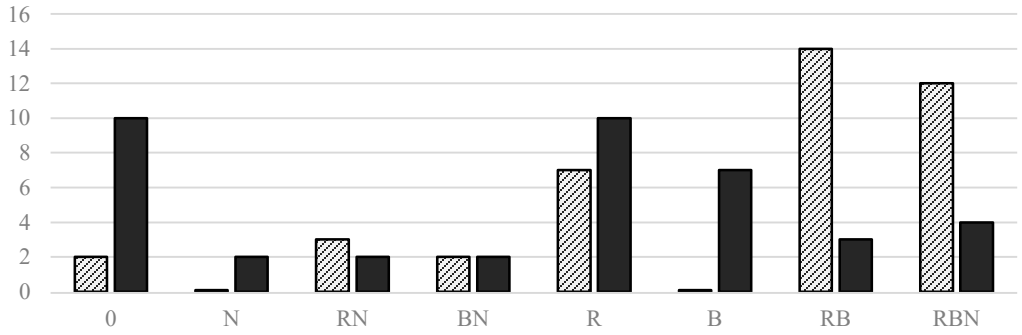


Figure 3.6: Visualization of the distribution of conjunctions of the three most prominent themes in the literature review: 'not constrained to reality' (R), 'better interaction' (B),'natural interaction' (N), and no statement (0). Explicit statements are shown in black, whereas explicit and implicit statements are hatched.

3.2.4 Results

To answer the motivating question **R1** for this literature review, the findings of the previous sections can be summarized as:

- In many cases, it is not clear what the term 'super-natural' explicitly refers to. Definitions and conceptual frameworks that could clarify what 'super-natural' explicitly means in each case and how it relates to other terms, e.g., [Ste17a; NB15; TJ01; SU94; NSN16; Gla19], are usually not referenced. Researchers usually describe what they mean when they employ this term instead of referencing a definition, or they provide related examples of super-natural interaction. In many cases, the term leaves room for interpretation. In the analyzed literature corpus, only two explicit definition proposals were identified. One by Steinicke [Ste17b] as well as a derived version by Lubos [Lub18], and one framework relating 'super-natural' to other terms by Nabiyouni and Bowman [NB15].
- Two main interpretations of the term 'super-natural' can be found: *super-natural abilities*, which are utilized by users to solve specific tasks (see Fig. 3.3), and *super-natural narratives*, which incorporate components in an IVE that cannot scientifically be explained (see Fig. 3.4). Some authors also refer to *super-natural environments*, which are often closely related to *super-natural abilities* by providing the conditions for super-natural interaction to occur.
- Super-natural abilities are a general concept that includes improved senses (e.g., visual or auditory), physical actions (e.g., locomotion and object manipulation), and changes to the user's virtual body (e.g., increasing size).
- For super-natural abilities, the mainly referred and often combined properties are (R) removing constraints of reality, (B) increasing power and (N) providing natural ways of interaction, whereas narrative elements (S) and effects on the feeling of presence (P) are only referred to in rare cases (see Fig. 3.5). However, no pattern for the inclusion of specific properties can be identified, and in all cases, specific properties are typically simply not mentioned, rather than being explicitly rejected.

- Many concurrent terms are used in a similar way to 'super-natural.' A clear delineation has not been identified in the literature review. The term 'super-natural' belongs to a group of terms ('superpowers,' 'superhuman,' 'beyond-real') that are not frequently employed to describe 'magic' interaction techniques (see Tab. 3.1). A marginal increase in the use of 'super-natural' is visible from the year 2014 on, but it can still be considered a relatively uncommon term (see Fig. 3.2).

3.2.5 Discussion

Overall, the term 'super-natural' can be considered a descriptive term that is sometimes used in research to emphasize that a specific approach for interaction design in VR does not have its roots in physical reality but instead follows the idea of creating novel ways for interaction that are imagined by VR designers, engineers, and researchers. The actual properties of 'super-natural' interaction are not uniformly presented, however, researchers use this term to describe interaction techniques that are based on shared principles, most prominently, rejecting limitations of reality, enhancing capabilities of users beyond human limitations, and, in some cases, providing simple ways for interaction. The relationship to other terms, such as 'magic,' 'empowerment,' and 'superpowers' remains an open question that cannot be answered with this review alone. Considering the usage of these similar terms in the research literature, it is likely that they are typically used as synonyms without conveying a specific meaning that differs from 'super-natural.' However, the findings produced by this literature review are limited, considering that, in most cases, there exists no explicit characterization of the term and related themes, which leaves room for interpretation. In the future, it would be beneficial to clarify the differences and reach a consensus on specific meanings to facilitate communication among researchers.

Furthermore, a clear distinction between 'super-natural' interaction techniques (which are researched in this thesis) and 'super-natural' narratives can be beneficial. In the performed literature review, both approaches to interaction have been described, with 40 publications referencing abilities and 23 publications referencing narratives. Both describe fundamentally different aspects of IVEs and interaction design. Depending on the definition of super-natural interaction techniques, narrative elements do not necessarily fit this classification, which, however, contradicts the position of Nabioyuni [NB15], in which the narrative story elements associated with the interaction are the central aspect to distinguish non-natural and super-natural techniques. Such story elements can be considered to form a different approach to interaction design that incorporates imaginative elements, such as gods, ghosts, and magic, including science-fiction content [RY18] into the design. This approach could also be referred to as 'supernatural' (without the hyphen, in the same way, the term is commonly spelled to describe paranormal events and beliefs in everyday language) or, to allow a better delineation, as 'fictitious.'

Although the use of the term 'super-natural' has increased since 2014 in research literature, this does not necessarily reflect a growing interest in this specific aspect of interaction. A plausible factor for this increase is the general availability of VR hardware following the release of the Oculus DK II in 2013, and subsequent systems, such as HTC Vive and Oculus CV in 2016. A higher availability of hardware also enables more researchers to investigate experimental systems and non-fundamental questions regarding VR, such as super-natural interaction. Finally, this literature review presents an overview of the use of the term 'super-natural' until December 31st, 2022. However, the thematic interpretation of the term did not change significantly between the start of this dissertation in 2018 and the literature review (see Fig. 3.5), so the point in time for conducting this research did, presumably, not have an impact on the presented analysis.

3.3 Working Definition: Super-Natural

The presented literature review forms the foundation for the proposition of a definition of super-natural interaction, which is further developed in the subsequent chapters of this thesis. The most reported themes in the literature review regarding super-natural abilities were: i) removing constraints of the real world (**R**), ii) natural and intuitive use (**N**), and iii) better ways of interaction compared to real-world interaction (**B**).

The combination of non-realistic (**R**) superhuman powers (**B**) has often been referred to as 'magic' interaction techniques [SU94; Shn03; BMR12; LJ+17; Ser+18]. In 1986, Smith defined magical features for *physical-world metaphors* in user interfaces as "those capabilities that violate the metaphor in order to provide enhanced functionality" [Smi86]. Smith used the term *violation of the metaphor* in the Alternate Reality Toolkit to describe the modification of real-world metaphorical physical properties such as appearance, physical forces, and position to allow new forms of interaction [Smi86]. An essential aspect of this description is that magic techniques not only reduce realism but also provide enhanced functionality as one key component. While this perspective was not a necessary requirement identified in the literature review, enhanced functionality is also included as a key component the approach for defining super-natural in this thesis to account for the etymology of the term. Spiegelberger emphasizes in his theological analysis of the term 'supernatural' that 'supernatural' "obviously means more than mere non-naturalness. It implies a certain superiority of the non-natural, as expressed in the prefix "super." Not everything excluded from the field of nature is automatically above nature [...]" [Spi51]. This position can also be applied to differentiate between super-natural and non-natural interaction techniques. Both describe design approaches that do not aim at recreating interaction found in the real world, however, the former implies some form of improvement, whereas the latter does not necessarily include this aspect.

Natural interaction (**N**), on the other hand, can be described as a style of interaction "whose conventions are easy to learn and whose commands are powerful, yet easy to use," [DF73, p. 24] following the description by Thomas De Fanti. In De Fanti's work, this statement regards language-like commands as the primary input for computer systems, with the system designed to provide a *habitable* environment. He describes that full *habitability* would be reached when users are interacting effortlessly by mere thought and imagination with a computer system without having to think about the interaction itself and not requiring any discrete specifications of interaction [DF73]. The programming language proposed in De Fanti's work aims to provide a user interface that maximizes habitability by allowing users to type commands as they would express their thoughts in a human-like dialogue without following a computer-determined syntax, which was, however, limited by the technological constraints of systems of that era. However, the fundamental concept of habitability can also be applied to physical interaction in IVEs, which are designed to require no or minimal additional effort from users to understand and apply the individual means of purposefully acting within the environment. This includes actions that are learned in the real world and transferred to the IVE, as well as VR-specific means of interaction that are designed in such a way that learning is facilitated.

The descriptions of magical features by Smith [Smi86] and natural user interfaces by De Fanti [DF73][18] can be generalized and combined to propose a definition of super-natural interaction techniques that is aligned with the most reported themes in the literature review:

> **Working definition: Super-Natural Interaction Technique**
> An interaction technique is called 'super-natural' when i) it reduces realism in order to provide enhanced functionality, and ii) it is easy to learn, powerful (in the sense of enabling users to accomplish a task efficiently), and easy to use.

Following this working definition, super-natural interaction techniques combine three requirements, the most frequently reported themes **R**, **B**, and **N** in the literature review, and encompass, colloquially expressed, all *magic interaction techniques that are natural to use*. This makes super-natural interaction techniques a subclass of magic interaction techniques instead of treating both terms as synonyms. They blend "concepts from the users' familiar reality, including already well-established digital concepts ..., and the expressive power of digital computation" [JRG14] to enable users to perform novel actions that, despite not being part of everyday life, feel as natural as everyday interaction. At its core, this proposed definition aligns well with the definitions by Steinicke [Ste17b] and Lubos [Lub18], aiming for a more generalized definition that builds upon existing definitions. In Steinicke's definition, natural user interfaces are based on naïve physics and the derived physics-based concepts (gravity, friction, elasticity, etc.) and sensorimotor skills (pushing, pulling, throwing, etc.), which provide a basic catalog for interaction concepts that facilitate the acquisition of derived skills [Ste17b]. In Lubos' definition, *natural human mechanisms* expand naïve physics to explicitly include complex motor skills, such as walking, and cultural skills, such as language. In the proposed definition in this thesis, this is further expanded to include all types of interactions that can easily be learned and used to solve a task, regardless of whether they were acquired in real-world interactions or exclusively in the IVE.

[18] The referenced works by Smith and De Fanti are the oldest identified sources to provide a clear and explicit description of what 'magic' or 'natural' means, even though the terms have been used more implicitly in research before.

CHAPTER 4

ENACTING VIRTUALITY

In the previous chapter 3, three main characteristics of super-natural interaction techniques have been identified. These three characteristics can be clustered into two aspects that are further conceptualized in this thesis: Super-natural interaction techniques, i) reject certain aspects of reality to enhance functionality, and ii) they are both powerful and easy to learn and use. In this chapter, the first aspect, reduced realism that enhances functionality, is analyzed using the enactive approach as a framework. IVEs enable the rejection of certain aspects of reality and may also introduce entirely new properties that cannot be experienced in reality. With an enactive perspective, this implies the construction of a distinct agent-environment system that has to be enacted by the user, which leads to the emergence of a system-specific co-determination and meaning.

In this regard, interaction techniques in IVEs can be considered to have an influence similar to the effect of using technology and its mediating effect on human-world relations in the real world. Therefore, this chapter first presents the philosophical concepts of i) *organ projection*, ii) *post-phenomenology*, and iii) *pattern theory of self*, and combines these ideas with enactivism to form a coherent framework that is suitable to describe the usage of super-natural interaction techniques in IVEs. The enactive perspective assumed in this chapter proposes that many forms of interaction in IVEs can be considered mainly embodied and subconscious and, therefore, do not require complex cognitive 'System 2' strategies (see section 2.3.2). Instead, they are conducted by enacting and executing acquired schemata[19]. This includes natural forms of interaction that correspond to actions in the real world, such as physical walking and the manual handling of objects as well as super-natural forms that involve the human body, such as teleportation techniques [Boz+16], flying [KL19], walking on walls [Gie+18] and non-realistic extensions of limbs [Pou+96b].

4.1 (Techno-)Philosophical Aspects

4.1.1 Organ Projection

Ernst Kapp, a philosopher of the industrial age, related technological development to the concept of *organ projection*, in which every invented tool is seen as a projection of humans into their environment that extends the biological limitations of the human body [Kap18]. With this perspective, the hammer becomes a more solid fist, axes are seen as better hands to cut wood, knives and scissors are sharper and more durable fingernails, and the telegraphy network corresponds to the communication in the nervous system [Kap18]. The human hand, as one of the main effectors in the interaction between humans and the physical world, has a threefold function in Kapp's philosophy: (1) as a natural way of interacting with the world, and hereby (2) as the prototype for many mechanical tools, and (3) as primary means of transforming components of the human body into physical tools [Kap18]. By analyzing the use and nature of tools, it becomes possible to analyze the nature of humans if it is considered that a "mechanism, which is unconsciously formed on the model of an organic prototypal image ... serves retroactively in its turn as the prototypal

[19] Schemata are described in more detail in chapter 5.

image through which the organism ... is later explained and understood" [Kap18, p. 24]. According to Kapp, the effects of using tools not only affect humans on the individual level but also society and culture as a whole: "[T]he first tool to proceed from the human hand was the actual impetus for the development of culture ..." [Kap18, p. 170].

Ideas similar to *organ projection* have been proposed by various authors. Building on Kapp's concept, Gehlen distinguished between *organ strengthening* of existing features, *facilitation* of the required effort to perform an action, and *replacement* of our natural features [Geh80][20]. The human is seen as a "deficient being" [GR88], which is not specialized for its environment and, therefore, fully dependent on the use of technology to ensure survival. However, the intellect of humans allows them a "world-openness" , that enables them to create a cultural second nature [GR88] that relieves them from "the necessity to undergo organic adaptations to which animals are subject, and conversely allows [them] to alter [their] original circumstance to suit [them]" [Geh80]. Bloch criticized in the context of industries and machines that, ultimately, the interaction with technology does not correspond to natural interaction between humans and their environment and becomes *de-organized*[21] [Blo+86]. Examples of de-organized tools are propeller-driven planes, which neither resemble the wings of birds nor any limbs of the human body, and mechanical sewing machines, which perform the task of sewing in a fundamentally different way compared to the natural way of employing human capabilities [Blo+86]. Freud described that, throughout history, humans have projected their desire for being omnipotent and in control of life onto the image of gods, and, using technology, they aim at reaching this state [Fre10]. He further described the image of a *prosthetic god*: "Man has become a god by means of artificial limbs, so to speak, quite magnificent when equipped with all his accessory organs; but they do not grow on him and they still give him trouble at times" [Fre10, p. 66]. In the field of Activity Theory, Leonitev introduced the concept of the *functional organ* that shares similarity to organ projection, [Nar98]. He described that every tool used by humans is related to our human body, and the use follows the goal of extending our natural abilities.

These fundamental concepts, *organ projection* and *de-organization*, are also evident in VR and HCI research in the form of developing, using, and imagining new means of interaction in an artificially created virtual world. Often, these interactions can be interpreted as "reflect[ing] the powers we would wish to have" [Bin00] in reality to control today's increasingly technology-infused world or even to provide completely novel ways of interacting. Enhanced senses, as 'passive' organ projections, enable new ways of meaningfully interacting with an altered environment. In contrast to other interactive technological media, such as desktop PCs and smartphones, VR research is, further, an interesting application for implementing organ projections, considering the immersive nature of IVEs and the embodiment of a freely modifiable avatar. For example, by modifying the 'real' organs and limbs of the user avatars, it becomes possible to increase the range of hand movement [Pou+96a; EWK18], or even enable users to interact with their environment using additional limbs [LBS14]. In this regard, 'avatar' refers not only to the graphical representation of the user but further incorporates other aspects of embodiment, such as a modified body schema as the foundation for interaction with the environment. IVEs enable designers and developers to create specific organ projections and related virtual tools that would not be possible with the constraints of reality. On the other hand, the separation between the use of VR technology and the real world is often more consciously present to users compared to other technology, such as watches, electric lights, and smartphones, as IVEs are actively engaged. Virtual organ projections in IVEs are, due to their limitation of

[20] The translation of the German terms is not uniform. The original terms are 'Organverstärkung' (strengthening), 'Organentlastung' (facilitation), and 'Organersatz' (replacement).

[21] The original German term used by Bloch is 'Entorganisierung'.

being usable for only a limited amount of time in an actively engaged VR system, less integrated into our everyday perception of reality than other technologies. A structured way of thinking about the changed relations between humans and their environmentis provided by incorporating post-phenomenological concepts into research.

4.1.2 Post-Phenomenology

Post-phenomenology extends the tradition of phenomenology, which focuses on the subjective first-person experience of the world [Bul13], by incorporating a broader perspective that assigns technology an active role in the mediation of human-world relations [Ver16] in complex and, often, subtle ways [Fra19]. It is further related to a post-modern [RV15] understanding of the specific challenges that emerge for individuals in the human society of the digital age. An important term that is adapted from the related school of phenomenology is 'intentionality', which describes the directedness of actions and perceptions of objects and the environment towards a subject-dependent goal [Mor18; Dil71]. In this view, objects do not exist independently of a person's ideas about them, and thinking about objects often implies thinking about how an object is used in a specific context; a conscious experience "is always consciousness-of-something" [Gal22, p. 53]. According to the phenomenologist Merleau-Ponty, our engagement with the world is rooted in the embodiment of the perceiver, and the perceiver's bodily experiences and sensory perceptions form the foundation for conscious experience [MP02]. According to him, "[t]he thing is inseparable from a person perceiving it, and can never be actually in itself because its articulations are those of our very existence, and because it stands at the other end of our gaze or at the terminus of a sensory exploration which invests it with humanity" [MP02, p. 373]. In this phenomenological view, intentionality emerges as a dynamic process wherein the *lived body*[22] serves as the mediator between our subjective intentions and the surrounding objective reality. A famous illustration of intentionality and the role of the body presented by Merleau-Ponty is the blind man who actively incorporates a walking stick as an extension to his body to perceive the world in a similar way to visual perception, rather than simply using the walking stick as a passive tool for support [MP02]. To a blind person, the walking "stick is no longer an object perceived ..., but an instrument with which he perceives. It is a bodily auxiliary, an extension of the bodily synthesis" [MP02, p. 176] that changes how a blind person approaches their environment, and how they think and act.

Post-Phenomenology expands the idea of intentionality by proposing that tools and technology created by humans provide an *instrumental* or *technological intentionality* [Ver01] to users. This leads to changes in the relation between humans and their environment as the "tool or equipment becomes a *means* of accomplishment" [Ihd17, p. 33], which either enables achieving goals or facilitates achieving them. In some cases, tools even produce new tasks by introducing new ways of perceiving the world or creating artifacts that differ fundamentally from previously known objects. For example, the development of optical technology (e.g., microscopes and astronomical telescopes), which enhances the capabilities of the human eye, led to novel knowledge and gave rise to new fields of research that have fundamentally changed our understanding of the relationship between humans and the world [Ihd17]. Instead of seeing the negative effects of technology on humans, post-phenomenology does not criticize it as alienating humans from the world, but rather as defining the human-world relation in a constituting way [Fra19]: "Human-world relations are practically enacted via technologies" [RV15]. Four important ways in which technologies can mediate our experience of the world are described as [Ihd17]:

[22] In phenomenology, the *lived body* describes the subjective dimensions of an individual's bodily experience, whereas the *living body* describes biological and physiological aspects of a person's physical existence [SH21b].

- **Embodiment: (I − Technology) → World**. Technology is a mediator between the user and the world. For an embodied relation, the technology becomes *transparent* [Ihd90], which describes the user as not interacting with the world indirectly using technology but rather directly through the usage of technology. The use of technology leads to an embodied "experiencing in a particular way by way of perceiving through such technologies and through the reflexive transformation of [one's] perceptual and body sense" [Ihd90, p. 72]. Embodied technology enables the user to benefit from "the transformation that the technology makes available. Only by using the technology is [the user's] bodily power enhanced and magnified by speed, through distance, or by any of the other ways in which technologies change [the user's] capacities. ... [Users] want the transformation that the technology allows, but [they] want it in such a way that [they are] basically unaware of its presence." [Ihd90, p. 75]
- **Hermeneutic: I → (Technology − World)**. In a hermeneutic relation, technology allows for new ways of perceiving the world without transparency. These technologies are referential to some aspect of the real world and provide a means of actively engaging with the information they carry. Instrument panels, for example, provide insight into the processes and properties of a moving vehicle by translating certain variables (e.g., speed and fuel amount) to gauge and dial values [Ihd17]. Tools can, in this context, also be described as "instrumental phenomenological variations" [Ihd17, p. 67] that provide new ways of obtaining knowledge about the world.
- **Alteriy: I → Technology − (−World)**. Humans can also engage with non-transparent technology, which is also not referential to other aspects of the world. In such a case, "technology may emerge as the foreground and focal quasi-other with which [users] momentarily engage" [Ihd90, p. 107]. The direct interaction with technology and artifacts provides new opportunities for interaction and achieving goals. Examples of direct engagement with technology are robots [Ihd17] and toys [Ihd90] as technological artifacts. In particular, desktop computer systems can be seen as examples of an alterity relation [Ihd90] in which humans actively engage with a technological object in the world with its own tasks, affordances, and challenges.
- **Background: I (Technology / World)**. In the background relation, technology remains out of focus and is, for most of the time, not actively engaged by users. It is *phenomenologically absent* [Ihd90], but it still influences the life of humans. Hereby, it forms, together with other technological and non-technological components, the lifeworld in which users act. Often, such systems are automatic or semiautomatic, for example, thermostats [Ihd90], or electric lights that disappear from our focus as soon as they are activated [Ihd17]. Typically, background technology only becomes apparent in our conscious perception when it fails to provide the expected background.

The relationship with technology is typically only perceived at a subconscious level, and a transition between modes is possible, often depending on the context. First, by incorporating different types of technological relations into the lifeworld, a *field composition* [RV15] is formed that directs the awareness of humans to highlight specific elements of the environment. In some cases, technology creates new environments for perception and hides other elements from the conscious experience. On the other hand, the use of familiar technology is described as *sedimentation*, in which the history of engaging with technology determines how humans approach their environment and what they expect from interaction by providing "the pre-perceptive context that enables our current perceptions to occur with immediate meaningfulness" [RV15, p. 25]. Another important concept in post-phenomenology is *multistability* [Ihd90], the versatile nature of technology and its varying manifestations of use [Ihd17; Ros17]. Technological devices possess an inherent ambiguity

that yields multiple interpretations depending on the observer's expectations and previous experiences. For technology and artifacts, there is often not a singular, objective interpretation. Instead, the user has a subjective and interpretive role that depends on the specific context. Technology can further be analyzed from two perspectives. First, it can be examined at the individual level as *microperception* with a focus on "what is immediate and focused bodily in actual seeing, hearing" [Ihd90, p. 29]. Furthermore, it can be studied at the broader level of human society and culture as *macroperception*, which analyzes the impact on society as "cultural, or hermeneutic perception" [Ihd90, p. 29].

Considering the large impact of computer systems on today's society, HCI provides an extensive field for postphenomenological research. Brey identifies the roles of "computer systems as both cognitive devices and simulation devices" that help humans in interpreting the world and, furthermore, create a world on their own in which humans act independently from the real world [Bre05]. As a multistable technological artifact, computer systems can change the relation between humans and the world in different ways, for example, as *embodied* tools (using a smartphone for a phone call), *alterity* artifacts (playing games on a console), *background* systems (controlling a smart home environment) and by providing *hermeneutic* visualizations (analyzing the weather forecast in the Internet). VR and AR systems are further interesting as subjects of research as they "generate or represent objects and environments that form an addition to the physical world" [Bre05]. Such technologies can change the relations of humans to both the physical world and the virtual world as *parallel relations* [RV15, pp. 22].

4.1.3 Pattern Theory of Self

An interesting perspective to describe the effects of interaction with technology and postphenomenological human-world relations is the formation of a *pattern of self* that is developed during everyday interaction as well as interaction with a specific technology. The general concept of the self has been a topic of interest in various disciplines, including philosophy, psychology, and neuroscience. The properties and constituents of the self, as well as fundamental issues such as questioning its very existence, remain debated. One perspective on the concept of a 'self' is to understand it not as a unified, autonomous, and independent construct but rather as a constantly changing entity that is shaped by the social, cultural, and environmental context in which it is embedded. Various authors have introduced terms to account for this context-dependent notion of the self, for example, "the cognitive self, the conceptual self, the contextualized self, the core self, the dialogic self, the ecological self, the embodied self [and others]" [Str99]. The *pattern theory of self* [Gal13; New18] proposes that "the self is a flexible entity which is a unity of characteristic features integrated as a pattern in a situation and then developed further," [New18] which is further "anchored in the body and which determines the body as the anchoring unit for self-conscious experiences" [New18]. Furthermore, "a certain pattern of characteristic features constitutes an individual self" [Gal13], which implies the existence of multiple selves as flexible and interacting units of conscious experience that produce the foundation for situated cognition and behavior during interaction with the world and technological devices. Newen describes three layers of the self: an *embodied self*, which integrates characteristic features into a minimal core self, a *long-term self-model*, which changes slowly over time, and a *short-term self-model*, which allows a quick adaptation to the currently experienced context [New18]. Gallagher presents a conceptual list of aspects that shape a self which contains *embodied*, *experiential*, *affective*, *intersubjective*, *cognitive*, *narrative*, *extended* (incorporating objects external to the individual), and *situated* aspects [Gal13].

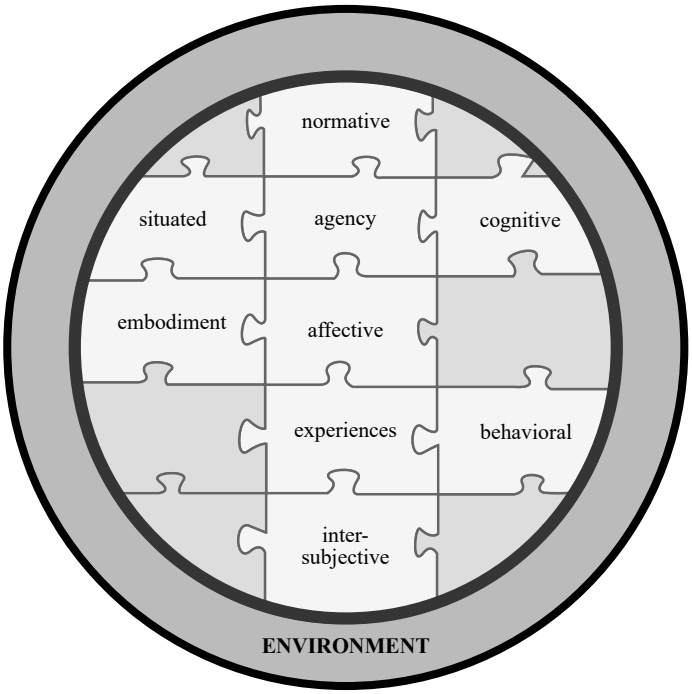


Figure 4.1: The pattern of self embedded in an environment. Important aspects (see [Gal13; New18]) are visualized as puzzle pieces. The model is not exhaustive, so some pieces are intentionally left blank.

Following this concept, it can be assumed that interaction with technology, such as IVEs, creates distinct technology-dependent patterns of self that can be visualized as a collection of characteristic constituents that form a self embedded into the social and technological environment (including extended aspects) (see Fig. 4.1). Experiencing a specific VR application leads to the creation of a distinct pattern of self that describes how the user perceives themselves (see Table 4.1) as part of the IVE. This pattern of self is the result of the artificially created environment, and single or multiple constituents may differ from other patterns of self. In many cases, the interaction in an IVE is specifically designed for a discrete application, and the possibility of applying this pattern is limited to this specific IVE. However, some applications share typical interaction techniques or transfer aspects of interaction from the real world. Often, changes in one constituent lead to changes in other related constituents. For example, the Go-Go technique presented by Poupyrev et al. [Pou+96a] can at first glance be considered to change the user's embodiment. However, this also affects agency in such a way that new potential actions are perceived, which influences the user's behavior, problem-solving strategies, and perception of the environment, ultimately leading to novel experiences that differ from reality.

The pattern of self that is created during interaction can be considered an emergent entity. Following the descriptions by de Haan, the technological medium that offers a technology-specific way of acting and cognition can be considered a *conjugate* that gives rise to a *type III*, *reflective emergence*, in which the observer, the user, is part of the very system that is emerging [Haa06]. A key characteristic of emergence is that the system forms an entity that cannot be fully described by the properties of its parts. Instead, it shows behavior that cannot be predicted from the underlying working principles. For the emergence of the technological pattern of self in VR, it can be assumed that multiple aspects have to be considered, which are, however, not clearly separated but interconnected and interdependent (see Table 4.1). The sum of these aspects forms a foundation for the conscious and subjective experience of acting within a technological medium.

Constituent	Description	Example
Cognition	Cognition describes both the adaptive sensorimotor coupling and sense-making, as well as rational problem-solving, e.g., a modified means-end analysis.	Considering different locomotion techniques to travel to a location, such as natural walking, flying, or teleportation.
Behavior	Behavior is the purposeful physical acting within an environment altered by novel means of acting.	A user utilizes a teleportation technique to make exploration of an IVE physically less demanding.
Agency	Novel ways of acting lead to new intentionality and consciousness about potential actions.	To reach a far object at a distance, the user's arm extends beyond its natural length using a Go-Go technique [Pou+96a].
Environment	New affordances are perceived, or ordinary constraints do not apply.	To cross a river, the user utilizes a point-&-teleport technique [Boz+16], the distant shore affords teleportation.
Embodiment	Changes regarding body image and body schema, and sensorimotor coupling to the environment	The user has four hands that can be used for object interaction [LBS14].
Experiences	New interactions lead to unique experiences and narratives.	A user flies above a virtual city model using an "Iron Man"-interface [Kru+16].
Affectivity	Emotional responses to experiences differ from reactions outside of the IVE.	In a VR game, the virtual death of an incorporated avatar is not a significant event.
Normativity	Experiences are evaluated using a different normative framework (e.g., social norms and perceived threats).	Users might engage more easily in behaviors they avoid in real life due to phobias.
Intersubjectivity	Social roles and behavior are affected by new agency.	Superpowers lead to prosocial behavior [RBB13].
Situatedness	Acting in VR depend on the activity (roles and goals) and on the sociocultural context.	The sociocultural identity of the user's lived reality can be replaced by incorporating a different avatar.

Table 4.1: Description of important constituents of the pattern of self in VR.

Beyond VR interaction, it can also be assumed that not only IVEs create such patterns of self. It can be proposed that engagement with technology and artifacts leads, in general, to the enactment and development of distinct short-term patterns of self that are based on the involved constituents of the technological user-environment system. In alignment with Newen's theory (see [New18]), the long-term self-model, or 'embodied self' (see Fig. 4.2), is used to describe the 'everyday self' using only sedimented and fully-embodied technology. Other patterns that are 'close' to the core self describe the use of everyday technology that is perceived by users as part of their typical lifeworld, which is typically interacted with in an automatic and subconscious way. Following the classification in post-phenomenology, examples of such technologies are fully embodied and *sedimented* artifacts, such as cars used to travel or smartphones used for telecommunication, as well as background technology, such as using electric lights in dark rooms or keys to unlock doors. Some patterns of self that are activated during engagement with technology can also be 'more distant' to the core self, for example, when technology is not frequently encountered or generally unfamiliar. When the technological environment changes, users

need to construct a new corresponding pattern of self or adopt and modify a pattern that has been created in previous interactions with similar technology. With this conceptual approach, the interaction with technology, such as a specific IVE, can be described as the creation or adoption of an appropriate pattern of self that encompasses the relevant constituents of the user-environment system. All technological engagements create specific models that influence one another and allow users to adapt between different modes of operating technological devices (see Fig. 4.2).

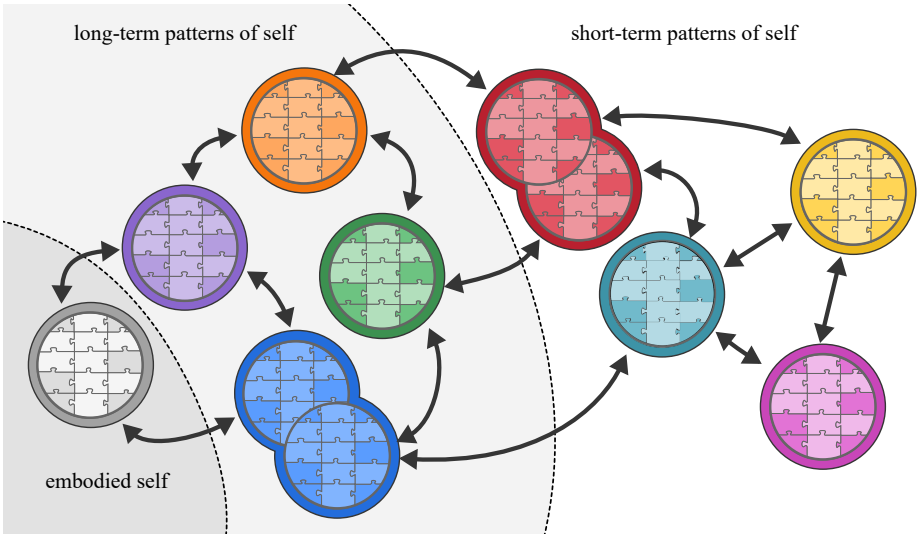


Figure 4.2: Transitions between the embodied self, long-term, and short-term patterns of self during interaction with different technology.

4.2 The Avatar-Virtuality System

4.2.1 VR Co-Determination

With the enactive approach, interaction within an IVE can be described as an embodied user avatar that engages with a virtual environment. Generally, interaction in VR systems often relies on whole-body input and can be considered more physical than interaction with traditional computer systems, such as desktop computer systems. IVEs provide a larger number of sensorimotor contingencies than traditional systems, which typically limit physical interaction to pressing buttons or the movement of a pointing device. Instead, the catalog of actions in an IVE can include typical physical actions found in the real world, such as natural walking or direct manipulation of objects, as well as novel actions based on physical movements. A common approach in VR research and applications is the re-creation of reality within an IVE [HHS19], which implies a correspondence (see Fig. 4.3) between both the user's avatar and the human user as an autopoietic organism (*subject correspondence*) interacting within the real environment, as well as between the displayed IVE and the real environment (*environment correspondence*). Correspondence can be analyzed at both a rational and objective level by comparing properties, and at a phenomenological level, regarding the subjective lived experience of interacting within an IVE.

Even though the avatar-virtuality VR system experienced in an IVE is embedded into the real world, it does not have to follow the ordinary rules of reality. The IVE is artificially created, so physical properties, actions, reactions, and meanings are, at first, determined by a designer with the possibility of creating novel or reality-contradicting interactions. Later, they are enacted by the user in a process of structural coupling of sensorimotor con-

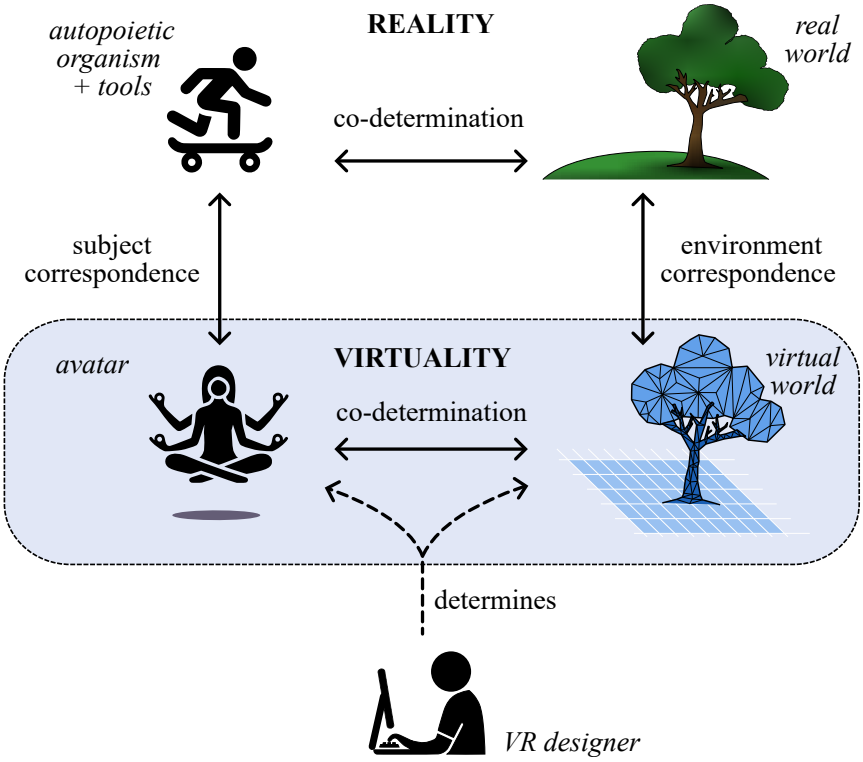


Figure 4.3: Correspondences and co-determinations between the autopoietic organism (including tools and other 'organ projections'), the avatar, the virtual world, and the real world.

tingencies and meanings. One strategy for reducing the effort required to enact the IVE is increasing the correspondence between reality and the IVE by transferring aspects of the real world to the IVE. A reduced realism can occur at the sensory and perceptual level of fidelity [MLP16], for example, a reduced rendering quality of stimuli, and at a higher plausibility level [Hof+20], e.g., reality-contradicting flying objects. Remarkably, users accept the rules of an environment with a low external plausibility as long as the displayed environment is coherent [Hof+20]. Both the avatar and the virtual world can differ from reality independently, thus changing the co-determination, which allows for an anthropocentric and an ecocentric perspective on the design of IVEs that are not necessarily coherent with reality. The ecocentric focus lies on aspects of the environment that are external to the user, such as violations of gravity, transparent objects, and other experiences that would be considered impossible in reality. Regarding the anthropocentric perspective, important sensorimotor contingencies, skills, meanings, and emotions are aspects that can be investigated, as well as the use of tools incorporated into the body schema. Typically, avatars in contemporary VR systems integrate only a subset of the sensorimotor contingencies that exist in human users, for example, head movement, which shifts the viewport for exploration of an environment, but no visible leg movement, as well as arm movements that change the spatial location of effectors (hands or controllers), but no haptic interaction with objects. Since skills are based on meaningfully coupled sensorimotor contingencies, facilitating the attunement of the user to the IVE is a primary aspect of creating usable IVEs.

In contrast to reality, which appears more stable regarding co-determination than only temporarily experienced IVEs, co-determination is not only an emergent property of the agent-environment system, but it also becomes dependent on the agent-environment-developer system. The co-determination is, on the one hand, determined by the VR developer who

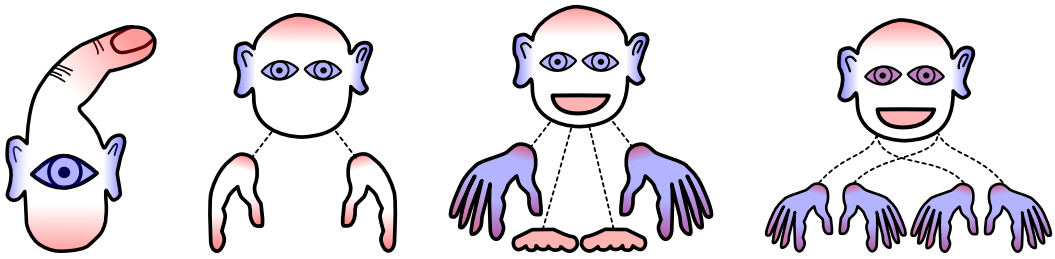
defines agent-world relations through implemented interaction techniques, and, on the other, an emergent system enacted by the user. The avatar-virtuality system can only be utilized in the intended way when the co-determination of the avatar and environment enacted by the user matches the intended determination of the developer. However, the co-determination of the avatar-virtuality system can also be enacted in diverse ways not intended by the developer as long as the user's sense-making and enacted meaning are relevant to some goals of the user.

It is challenging to describe the conscious experience of acting as a virtual avatar within a specific VR agent-environment system with properties that differ from reality. In his influential essay, "What Is It Like to Be a Bat?" [Nag80], Thomas Nagel critiques the reductionist approach to explaining mental states and consciousness through mechanistic principles. In his view, the unique properties of agents, in the essay's example, bats, form the basis for a subjective experience of the world that cannot be fully understood by agents with different properties, such as humans. An attempt to describe consciousness through objective factors is, therefore, not possible and cannot account for the subjective experience of an arbitrarily defined agent-environment system. In the same way, it can be assumed that a pure objectivist view on IVEs does not fully explain the effects of embodying an arbitrarily shaped avatar on the subjective experience in a world that does not follow the rules of reality. For example, teleporting in a non-Euclidean space in VR, in which the geometry of the environment may not follow the rules of the physical world, offers a subjective experience that cannot be directly mapped to our real-world understanding of space. The user's conscious experience in this environment would have a unique subjective character that may differ drastically from typical physical interactions.

4.2.2 Sensorimotor Contingencies & Embodiment

As one of the main aspects of virtual user-environment systems, the embodied virtual avatar with unique physical and quasi-physical features plays an important role. As mentioned in the previous section, sensorimotor contingencies are a key component that distinguishes VR from other types of media. A popular visualization of sensorimotor contingencies are the sensory and the motor homunculi, distorted humanoid figures with body parts scaled according to corresponding proportions of brain tissue in the sensory and motor cortex [PB37]. In HCI literature, another kind of homunculus can be encountered, which is not based on brain structure but on the involvement of components of the human body during interaction with a computer interface. O'Sullivan's and Igoe's "human being as seen through the computer's input devices" [OI04] (see Fig. 4.4a) depicts the input and output of a user during interaction with a graphical user interface (GUI). The system input is limited to moving a mouse pointer and pressing buttons, which can be achieved with a single finger. The system output is visually perceived in 2D or aurally in stereo sound. The interface homunculus visually depicts the constraints imposed on human users in HCI, specifically regarding limitations on input and output mechanisms. From an enactivist perspective, this depiction corresponds to the illustration of sensorimotor contingencies, which are employed during interaction with the computer system.

In VR research, sensorimotor contingencies have been described as an objective property of a system to provide immersion [Sla09]. In this context, sensorimotor contingencies include *sensorimotor actions* and *effectual actions* that influence the environment [Sla09]. With an enactivist view, both types are fundamental isomorphic actions that are grounded in the interaction with the real world through our human body, which are integrated into the IVE. Non-isomorphic actions, such as super-natural interaction techniques, are still dependent on certain human sensorimotor contingencies, however, these are not necessarily the same as in reality, or they may not even exist there at all.



(a) GUI (b) Basic VR (c) Advanced VR (d) Super-Natural VR

Figure 4.4: Four interface homunculi with highlighted sensorimotor contingencies (blue = sensor, red = effector, violet = both). The outline of (a) as well as the style of the other homunculi is adapted from [OI04]. Reprint from [DGS23].

In contemporary IVEs, the interface homunculus acquires more features and appears more human compared to the GUI homunculus.[OI04] In the following, we consider three types of homunculi for IVEs: One for basic VR systems, which encompasses IVEs with just one head-mounted display (HMD) and a pair of controllers as input devices, one for advanced VR systems, which represents possible capabilities of current-generation HMDs, and one for super-natural VR systems. For many basic IVEs (see Fig. 4.4b), instead of one finger, the homunculus possesses two hands, which are each tracked with 6 degrees of freedom (DOF) in 3D space. Each hand has only two fingers, one thumb, and one index finger, which are used to press buttons and triggers, or control the touchpad and thumbsticks of typical VR controllers. The head is tracked in 6 DOF and spatially related to the hands. The basic VR homunculus perceives audio in stereo and has stereoscopic vision. In advanced VR systems, the homunculus (see Fig. 4.4c) can make use of more sensorimotor contingencies. Instead of two fingers, data gloves and camera-based systems enable the tracking of all five fingers of each hand. Controllers simulate haptic feedback using vibrations, or complex devices are used to provide a realistic haptic sensation. Voice input can be used for commands and text entry, and the movement of feet can be captured with the aid of additional sensors. Some IVEs allow the utilization of novel sensorimotor contingencies for engaging with the IVE. For example, in quad-manual user interfaces [LBS14] (see Fig. 4.4d), the user controls an avatar with two pairs of hands and switches between hands by gazing at the pair of hands they intend to use. The user's eyes not only perceive the virtual environment but can also be utilized as additional system input, which implements new sensorimotor contingencies that are not present in the real world.

Changing the body of the virtual avatar determines the sensorimotor contingencies and, following the enactive approach, this implies altered ways of cognition (as defined in enactivism as regulated sensorimotor coupling with the environment) and, furthermore, behavior and skills. For the conscious entity, in this case, a human user, "real space is wherever perception and embodied action occur" [Cla04, p. 17]. Acting with a virtual body ultimately leads to changes in both our body schema (subconsciously used during physical actions) and body image (how the physical body is consciously perceived) [Gal86].

> "If perception is in part constituted by our possession and exercise of bodily skills ... then it may also depend on our possession of the sorts of bodies that can encompass those skills, for only a creature with such a body could have those skills. To perceive like us it follows that you must have a body like ours." [Noë04]

The altered sensorimotor contingencies also influence embodiment in a virtual environment, the "possessing and acting through a physical manifestation in the world" [Dou04, p. 100] in real time and real space [Dou04]. Dreyfus further identifies three interpretations

of 'embodiment' in Merleau-Ponty's *Phenomenlogy of Perception*: i) the actual body with a defined number and shape of limbs, ii) physical skills and response patterns that are developed in interactions with the environment and iii) cultural skills and understanding as a result of being embedded into a cultural world [Dou04; Dre96]. Ziemke identifies six different but non-exclusive notions of 'embodiment' in research that partly overlap with Dreyfus' three interpretations: "(1) structural coupling between agent and environment, (2) historical embodiment as a result of a history of agent-environment interaction, (3) physical embodiment as a physical existence within a physical environment, (4) 'organismoid' embodiment, the organism-like bodily form, (5) organismic embodiment of autopoietic, living systems, and finally (6) social embodiment" [Zie03]. All these interpretations are relevant for researching phenomena regarding embodiment, with some directly depending on the implementation in specific VR applications (the shape of the avatar's body and physical skills in the form of continuous structural coupling) and some (social embodiment and the cultural world) visible in the form of conventions and specific implementations in multi-user applications. Importantly, this enactive view is not limited to objective performance measurements and includes research on subjective and phenomenological effects.

This notion of embodiment, beyond realism and matching of appearance, contrasts with the sense of embodiment as defined by Kilteni et al. [KGS12]. In their definition, Kilteni et al. explicitly define the sense of embodiment towards an artificial virtual body as a quale that emerges when the properties of the virtual body are processed in the same way the properties of the user's biological body are processed [KGS12]. From an enactivist perspective, the sense of embodiment does not necessarily depend on the alignment of the virtual body with the real biological body of the users but instead can be related to the attunement of the agent to the structural coupling of the system with all included sensorimotor contingencies, both realistic and non-isomorphic. This form of embodiment may encompass both the body image (how we perceive our body) and the body schema (how we interact with our body).

4.2.3 Affordances & Agency

New properties of the agent or the environment create new affordances that emerge from repeatedly experiencing the interactions within an agent-environment system and deriving concepts for potential actions. The three central claims of ecological psychology as the domain of which the concept of affordance originates are: i) perception is direct and non-mediated, so it does not rely on inferring information from internal representations, ii) perception is tightly coupled to the formation and execution of action, and iii) perception aims at supporting the interaction with affordances (as a result of direct and action-focused perception) [Che11]. Gibson characterizes the concept of affordance as " neither an objective property nor a subjective property; or it is both if you like. ... It is equally a fact of the environment and a fact of behavior. It is both physical and psychical, yet neither. An affordance points both ways, to the environment and to the observer" [Gib14, p. 121]. From an enactivist perspective, an affordance is always related to an agent-specific niche in which a coupling between sensorimotor contingencies and environmental features is possible [Che11]. In the same way, Weiser and Brown describe the concept of affordance in the context of HCI as "a relationship between an object in the world and the intentions, perceptions, and capabilities of a person." [WB96] Stoffregen formally defines that an affordance exists neither as part of the environment nor as part of an agent, but instead emerges from the combination of properties of the agent and the environment in the agent-environment system.

> "Let W_{pq} (e.g., a person-climbing-stairs system) = (X_p, Z_q) be composed of different things Z (e.g., person) and X (e.g., stairs).
>
> Let p be a property of X and q be a property of Z.
>
> The relation between p and q, p/q, defines a higher order property (i.e., a property of the animal–environment system), h.
>
> Then h is said to be an affordance of W_{pq} if and only if
>
> (i) W_{pq} = (X_p, Z_q) possesses h
>
> (ii) Neither Z nor X possesses h." [Sto18]

Following the proposed working definition of super-natural interaction in chapter 3.3, one required property of a super-natural affordance is the implementation of some form of reduced realism that is enacted from specific non-realistic properties of the avatar-virtuality system. Formally, let $\mathbb{Q}_r$ be the set of all properties of a real person Z and $\mathbb{P}_r$ the set of all properties of the real-world environment X. For a super-natural affordance s, either $q \notin \mathbb{Q}_r$, or (non-exclusively) $p \notin \mathbb{P}_r$ for a specific configuration W_{pq}.

This can be visualized using three common teleportation techniques in VR: Point-&-teleport techniques [Boz+16] assign the capability to be able to initiate a teleportation event using gestures or controller inputs as a property to the user. In many implementations, the virtual ground affords standing and positional changes similar to the real ground, which makes teleportation solely a property of the user's virtual avatar ($q \notin \mathbb{Q}_r$). Orb teleportation techniques [HL19] using static orbs distributed in the environment, on the other hand, add an object to the environment with the property of transporting the user to a different location. When interaction with these objects is based on natural modes of interaction, for example, reaching for the orb and pulling towards one's head, the properties of the user's avatar and the real human body are equal. In contrast, the environment obtains new properties that are not present in the real environment ($p \notin \mathbb{P}_r$). However, when point-&-teleport techniques are limited to special locations that afford teleportation, both the user and the virtual environment have non-realistic properties ($q \notin \mathbb{Q}_r$ and $p \notin \mathbb{P}_r$). Furthermore, in some cases, it is not possible to clearly distinguish if the properties belong to the user or the environment; for example, in walking-on-walls experiments [Gie+18; Bec+19], the walking on walls can either be interpreted as an additional capability of the user in a realistic environment (the ability to walk on walls), or as a shift in gravitational direction.

It can be assumed that affordances in VR that are introduced by super-natural interaction techniques form unique *conjugates* [WS20] by combining different lower-level affordances in complex ways. For example, a point-&-teleport locomotion technique [Boz+16] combines other nested affordances, such as pressing a specific button or pointing at a desired target location. The affordance of the ground to be used for teleportation is, however, not the result of a combination of simple affordances, but instead, in the tradition of Gibsonian ecological psychology, a directly perceived *complex particular* [WS20]. The affordances in an agent-environment system emerging in this way imply new types of agency. Here, 'agency' not only refers to the sense of being in control of actions [KGS12], or possessing a 'power to act' [Kim20], which are typical interpretations in HCI research, but encompasses four *core properties of human agency* as defined by Bandura [Ban06]:

(1) *Intentionality*, the goal-directedness of actions.

(2) *Forethought*, the visualization of the future as a motivating factor for action.

(3) *Self-reactiveness*, the construction of plans and actions to pursue goals.

(4) *Self-reflectiveness*, the metacognitive ability to judge one's own actions and thoughts.

Introducing super-natural interaction leads to changes regarding these aspects of agency. Users can perform novel tasks, aim at new goals, plan different sequences of actions, and the reflection on actions taken may lead to other judgments of encountered situations and problems. Within the technological medium of VR, super-natural interaction techniques constitute novel means of achieving goals that are contrasted with traditional actions, e.g., flying versus walking, as a form of 'functional anti-realism' that is beneficial to the user in some way.

4.2.4 Congruence

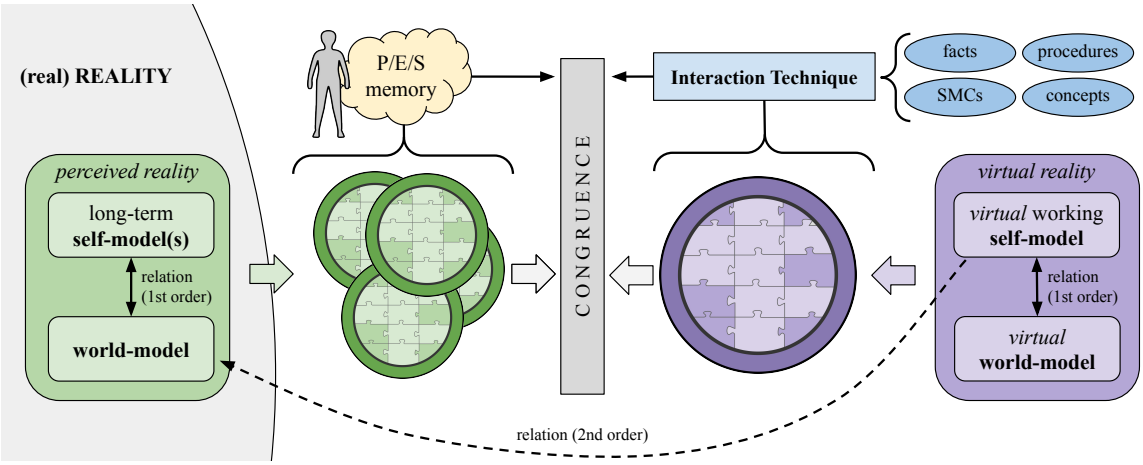


Figure 4.5: Congruence as a central concept describing the alignment of well-established long-term patterns of self and the virtual short-term self-model. The technology-mediated capabilities of the short-term VR pattern of self determine the relation to the virtual world (1st order relation) and possibly to the real world (2nd order relation).

Following the three presented philosophical concepts, a conceptual model can be created (see Fig. 4.5) for analyzing how well the user's long-term self-models, which are employed in the real world, and the short-term virtual working self-model, which is adopted during interaction with IVEs, are matched. It can be assumed that individuals possess a number of long-term models of the real environment in which they experience everyday life as autonomous agents, which capture the fraction of the 'real' reality [WHGA20] that is perceivable by humans as *perceived reality*. In addition to these long-term world models, users in an IVE possess a virtual world model that describes the currently experienced avatar-environment system with an IVE-specific pattern of self. Regarding realism, this model can range from a faithful replication of reality to modifications of certain aspects of reality to completely new alternative environments. The self-model and the world-model are constructed and constantly refined by interactions between the individual and the environment [New18]. When the pattern created during interaction with the IVE matches the long-term patterns of self in reality well, it can be called congruent.

Congruence is a central component that is investigated in this thesis. When a VR pattern of self is well-matched to existing long-term patterns of self, for example, simulating driving a virtual car that utilizes real-world means of control, this VR pattern of self has a high subjective congruence. On the other hand, other approaches that do not replicate real-world actions, such as a flying carpet, have only a low congruence for this specific user. Passive aspects, such as observable facts about the environment and the user's body, as well as memories and experiences, can objectively be analyzed regarding congruence (which can be called 'experiential congruence'). Furthermore, active aspects regarding motivations, problem-solving, goal-directedness, and meaningful physical interaction with

an environment (referred to as 'agential congruence') play an important role in interaction within IVEs. Social aspects (this can be called 'social congruence') can also be important, however, these are not further analyzed in this thesis. At its core, the concept of congruence in technology-mediated interaction shares similarities with what has been introduced as 'congruence' by the psychotherapist Carl Rogers. Rogers used the term 'congruence' to describe a genuine behavior and expression of feelings of therapists towards their clients that correspond to their true identity without introducing a mediating facade in the therapist-client relationship [WP20]. It refers to the alignment or consistency between a person's self-concept and their actual experiences and behavior. According to his model, therapists exhibit high congruence when their actual incorporated self and their ideal self align. In this thesis, this idea is transferred to the alignment of the incorporated technology-mediated self, e.g., incorporating an avatar and acting in a specific VR simulation, with long-term patterns of technology-mediated selves, such as riding a bike or using a mobile phone.

By entering an IVE, users incorporate a virtual working self-model (see also [Met18]) with specific abilities and properties that depend on the IVE. This virtual working self-model itself is not a static entity but dynamic, and users are required to constantly update their adopted pattern of self, which contains the virtual working self-model and world-model, based on the evolving situation in the virtual world, to stay in a state of attunement to the IVE. The adoption of an IVE pattern of self further exhibits a temporal dimension, as users can usually switch between different IVEs with unique interaction techniques and environmental behavior instantly, which necessitates quick adjustments to the adopted patterns, or, if the tensions between application-specific patterns are too high, incorporation or construction of a new, better-suited pattern of self.

Working definition: Congruence
Congruence describes the similarity between a specific technology- and interaction-dependent short-term pattern of self and the established user-specific long-term pattern of self.

Particularly when the IVE does not aim to recreate reality but instead introduces super-natural interaction techniques, the formation of the interaction- and environment-specific virtual working self-model becomes an interesting topic for analysis. Super-natural techniques often require users to adapt or develop a unique short-term pattern of self in response to the virtual environment and the provided means of interaction, for example, a self that can teleport to any location [Boz+16], or a self that possesses an increased range of arm movement [Pou+96a]. Embodying this IVE-specific avatar is a critical aspect of the virtual working self-model that introduces an additional layer of complexity when the virtual avatar differs from the real-world agent, considering appearance, capabilities, and actions. The employed interaction techniques encompass means of input and output, as well as mappings between actions and system feedback. The specific means of interaction can show similarities to previously enacted interactions in our procedural ("how?") and episodic and semantic ("what?") memory ('P/E/S memory' in Fig. 4.5).

When users first experience an IVE that provides an IVE-specific virtual pattern of self that differs from their long-term patterns of self, they perceive a tension that needs to be resolved, typically by enacting and training IVE-specific skills to facilitate a rapid transition between different patterns of self. For example, point-&-teleport techniques [Boz+16] as well as orb and portal teleportation techniques [HL19] are very common locomotion techniques in IVEs, which are, during the first experience, vastly different from real-world locomotion. Their design allows for a fast enacting and mastery of these skills, which reduces the tension in such a way that experienced users are enabled to employ these

teleportation techniques as natural means of locomotion in IVEs, even though certain aspects of patterns of self, such as, *agency*, *embodiment* and *experiences* remain different from real-world locomotion. In such a case, the interaction technique can be highly usable, but the related pattern of self maintains a low congruence with patterns of reality.

Congruence is highly subjective and related to the individual's prior experiences. For example, the experienced congruence of flying in an IVE is presumably largely dependent on the included narrative and story elements. Flying a magic carpet in an IVE would produce a low congruence for most users regarding prior experiences, as it includes a clearly magical device that is not encountered in reality. In a similar way to a history of real encounters with a specific artifact and the subsequent knowledge about the use, a fictional narrative may support users in identifying the fundamental application of a specific super-natural interaction technique. For example, knowing that 'a magic carpet can be used for flying' helps the user understand what a virtual magic carpet is used for, even though the involved procedural knowledge does not exist during the first use. In contrast, flying a virtual airplane would not contradict reality for most members of today's population (high experiential congruence) but still change the perceived abilities of most users (low agential congruence). However, those users who already possess a distinct short-term self-model for flying an airplane, such as pilots, would perceive the increase in abilities less prominently. The perception of controlling a magic carpet or an airplane is mainly dependent on the implemented means of control and interaction design. In many cases, a natural, intuitive, and easy way of interacting is beneficial and aimed for during development. In the case of an airplane, this may lead to the design decision of implementing a simplified interaction that is easier to learn than flying a real airplane. In such a case, the usability for applying this form of locomotion is reduced for novice users by minimizing the time required to learn how to control the system (lower tension between the interaction techniques and existing procedural knowledge). In contrast, the experiential congruence is reduced for pilots who experience contradictions regarding interaction in the IVE and in the real world.

4.2.5 Enhancement

'Enhanced functionality' is a key property of the proposed working definition of super-natural interaction (see chapter 3.3). In the context of super-natural interaction in VR, the parallel relations [RV15] regarding VR and the real world as two distinct types of relations need to be distinguished. Enhanced functionality can regard the real world as an external frame of reference, as well as the virtual world. This can be referred to as a *first-order relation*, in which the relation between the user and the virtual environment is analyzed, and as *second-order relation* in which the focus is directed at the interaction between the user and the real world through a VR or AR system (see Fig. 4.5).

In the first-order relation, the IVE provides a technological medium in which different virtual tools and interaction techniques change the relationship between users and the virtual world. It can be assumed that these changes are perceived by the user in the same way as if the employed tools and interaction techniques could be used outside of the IVE in the real world. Virtual objects in IVEs are, from a post-phenomenological perspective, not necessarily perceived as fundamentally different from technological artifacts in the real world, and new post-phenomenological relations (e.g., hermeneutic, embodied, alterity, and background) follow the same principles as technology in the real world. The role of employed tools and interaction techniques in IVEs can, therefore, be interpreted as corresponding to the role of technology in the real world. For example, teleportation as a super-natural locomotion technique (e.g., as point-&-teleport [Boz+16]) can be analyzed as either a transparent *embodied* relation in which the enhanced capabilities are expected and naturally employed by users, after being used sufficiently often, forming a VR sedi-

mentation. However, depending on implementation, teleportation can also be described as an alterity relation when users have to actively select a magic tool that provides the means of locomotion (e.g., orb teleportation [HL19]). Other aspects of IVEs may be interpreted as background (e.g., a virtual floor on which the user 'stands,' or simulated autonomous agents) or as hermeneutic (e.g., visualizing the number of targets hits in a virtual archery setting within the IVE). During the interaction, different aspects of VR and AR systems may transition to a background relation, for example, the controllers in hands become, with more proficiency in use, phenomenologically 'transparent' and are not perceived as the focused objects of interaction.

In contrast to the internal first-order relation, the second-order relation analyzes how VR and AR technology change the relation between humans and the real world. First, these systems provide an alterity relation in which users actively engage with these systems. This relation is also maintained in applications in which users only interact with a virtual environment, for example, in single-player games. However, whenever these systems are not isolated from the real world, the relation between users and the real world may also change. Collaborative IVEs and multi-user environments, for example, provide new means of communication for both professional tasks and leisure activities. With a good design and sufficient features, the IVE technology can become 'transparent,' and users may interact freely through the means provided by the VR or AR system. IVEs can also provide insight into the real world, for example, by allowing users to explore a virtual 3D globe or virtually visit different locations. Transferring aspects of the real world to the IVEs in an augmented virtuality system or transferring informative content to an AR display leads to a hermeneutic relation in which the IVE provides new ways of understanding the real world. In the future, with smaller and more affordable devices, VR, MR, and AR technology may even possess the potential to become fully embodied by providing well-integrated technological solutions into our everyday lives. In both the optimistic and pessimistic vision, this may even lead to the proposed *cyborg relation* [RV15] in which technology and humans form an inseparable unity.

Working definition: Enhancement
Enhancement describes a change in the user-environment relation that is beneficial to the user. *First-order enhancement* limits reference interaction techniques to closely related patterns of self within the same technological medium (e.g., the real world, or a specific virtual environment), whereas *second-order enhancement* references an interaction outside of the technological medium (e.g., a VR interaction in comparison to a real-world action).

The motivation for creating a super-natural interaction and inherent acceptance of created tensions can, in many cases, be interpreted as a form of organ projection that is based on the challenges we perceive in our individually perceived everyday lifeworld. As presented in section 4.1.1, the use of technology has changed the relation between humans and the world throughout history to provide a *second nature* that alters the environment as a beneficial foundation to human life. Embedded into this *second nature*, VR systems provide new opportunities that, in contrast to other technology, can be fully detached from the constraints of reality. One main motivation can be to provide productive applications that benefit from super-natural interaction by reducing the challenges of work, especially with digital content, e.g., modeling 3D objects in VR or by providing new ways for real-time communication. In this context, the different modes of organ projection proposed by Gehlen and Block, *strengthening*, *facilitating*, and *de-organization/replacement*, can further be used to describe the experienced changes of *agency*, for example, making tasks

easier or enabling users to perform tasks and procedures that are not possible in the real world. In contrast to productive environments, it is also possible to provide interesting and entertaining environments that aim at individual preferences and leisure to promote creativity, unique experiences, and personal enjoyment, with super-natural interaction as an enabling approach for interaction design. In all cases, the new user-environment systems are intended to have a positive effect on the user's meaning of interaction, such as increased task performance in regard to a reference technique, or positive subjective effects.

CHAPTER 5

SCHEMATA-BASED INTERACTION

In this chapter, the second property of the presented definition, "powerful and easy to learn and use," is investigated. The goal of this section is to provide a framework for understanding and analyzing embodied interaction in IVEs using the enactive approach as a theoretical and conceptual foundation. In this chapter, the enactment of interaction schemata as fundamental representations of knowledge about system interaction and their relationship to enactivism are analyzed as a 'bottom-up' approach. The enactive approach offers an interesting perspective on conceptual aspects of 3DUIs in VR. While current research regarding the application of enactivism to VR is mostly conducted from a philosophical perspective (see, e.g., [Can22; RVF22; CS14]), this section aims further at transforming theoretical aspects of enactivism (namely, *co-determination*, *sensorimotor contingencies*, and *sensorimotor schemata*) into tools which can be used in practice to analyze interaction.

5.1 Building Blocks of Embodied Interaction

The function of computer systems is based on information processing, the manipulation and interpretation of symbols expressed in a binary language. This underlying structure can be recognized in CLIs, where word tokens correspond to code execution and parameters. In the case of 3DUIs, however, this underlying working mechanism can be considered neither visible nor relevant for users during interaction. 3DUI interaction is, from a phenomenological perspective, much more related to how we experience and interact with the real world rather than an exchange of symbolic information. The functions of artifacts and actions "are not determined by the inner workings of artifacts, if any, but by the purpose assigned to them by designers and users" [Bre05]. There is a tension "between the way in which we communicate experience and knowledge through language (the structure of logic, composition, reasoning, symbolization, etc.) and the way in which we exercise, experience, and live life as an embodied practice" [DPBB17, p. 12] that presumably also expands to the physical interaction between users and computer systems, especially in the context of VR interaction. The encoding of interaction into schemata that "turn action into meaning" [Dou04, p. 183] accounts for this embodied practice, which is based on implicit, experience-based knowledge rather than explicit, symbolic interaction. Physically interacting in 3DUIs, such as navigating an environment or manipulating objects, is often based on subconscious and automated implicit knowledge, or, as Di Paolo describes it: "the body knows" [DPBB17, pp. 12].

Different skills show varying levels of cognitive and sensorimotor influence, but all actions, even cognitive operations, are ultimately expressed physically with the user's hands or other parts of the body as "the medium through which the cognitive activity gains expression" [CNM83, p. 358]. In the case of CLIs, the physical actions correspond to typing on a keyboard to provide symbolic input to the computer system. In contrast, 3DUIs can be considered embodied interfaces, often involving a larger number of body parts and more expressive movements. This physical interaction in 3DUIs can often be better explained as a closed-loop cybernetic system in which a self-organized agent continuously acts upon an environment using embodied schemata in response to familiar situations or problems,

instead of an agent forming and manipulating a mental representation of the observed external world. In many cases, the physical actions are not directly linked to conscious cognitive processes and operate at a lower level of processing as *pre-intentional acts* and *action-oriented representations* [Gal17].

Pre-intentional acts are motor actions that are targeted towards the agent's intention, such as moving hands and fingers in the right position to grab a ball [Gal08], that are not consciously controlled but contribute to the agent's activity. Rowland describes pre-intentional acts as *deeds* that form a middle ground between intentional *actions* and sub-intentional *doings* [Row11]. In a strict sense, intentional *actions* are subject-related teleological processes that can be consciously described. This intentionality during acting "reaches down into the motor elements that serve the intentional action" [Gal08]. At the lower end are *doings* that are free of intentionality and "insulated from rational considerations and assessments" [Row11, p. 110]. They form an embodied skill that can be thought of as an action-oriented representation that encompasses a single action that can possibly be performed by the agent in its specific context [Gal08], which receives its meaning by being implemented in an intentional action. Action-oriented representations are i) action-specific (aimed at particular behavior and possible actions within the environment), ii) egocentric (spatially related to the agent's coordinate system and point of perception), and iii) intrinsically context-dependent (situation-specific coupling between the agent and environment that defines the possible actions in the context of the activity) [CW12]. Conscious and intentional actions, which are typically consciously executed, often invoke a network of "motor control processes and [pre-intentional acts] that contribute to the accomplishment of actions" [Gal17, p. 104] which are carried out on the subpersonal level [Gal17]. Action-oriented representations differ from symbolic representations used in information processing theory and correspond to sensorimotor schemata that are constructed as a result of a history of couplings between the agent and the environment, as well as repeated experiences of action-perception loops. Garbarini writes:

> "There is no construction of a symbolic representation, but there is representation and, with it, a form of constructivism: the sensorimotor scheme, inasmuch as a scheme always has an anticipatory component, is in itself a mental representation in which the experience is ''constructed'' on the bases of categories, which are not longer theoretical, but pragmatic, deriving from the dynamic interaction of the organism with its adaptive environment." [GA04]

Humans can be viewed as agents that possess a highly complex open sensorimotor agency[23] that gives them the capacity to adapt and learn new sensorimotor schemata that can be integrated into the overall network of related sensorimotor schemata [DPBB17]. For complex interaction, sense-making can rely on the incremental enacting of smaller portions and single steps of the interaction to allow for the construction of a *scaffolding* as a supporting structure that integrates learning and doing in a seamless process [WW11]. The most fundamental units are *motor schemata*, which encompass the knowledge about the physical performance of actions, and *perceptual schemata*, which facilitate the recognition and interpretation of observations [Arb92]. To some degree, this also applies to certain cognitive tasks and perceptive skills in such a way that "[p]roblem-solving behavior will, with practice, become cognitive skill" [CNM83]. For example, chess playing is considered to be largely based on pattern recognition [Chi+81; Dre96] and applying the correct knowledge to solve a specific "problem schema" [Chi+81].

[23] Di Paolo proposes five levels of agency with *open sensorimotor agency* as highest level that provides agents with "the adaptive capacity to learn new sensorimotor schemes in an open manner and integrate them in the overall network" [DPBB17, p. 170].

HCI relies on both conscious information processing for System 2 cognition and automated physical interactions for System 1 interactions. Both systems are interlinked, and interaction techniques can often be considered a combination of declarative knowledge (facts about the system), conceptual knowledge (image schemata and metaphors), and procedural knowledge (physical interactions), with varying influence depending on the task and the interface. For 3DUIs and spatial interaction, the employed procedural knowledge and subconscious application of bodily sensorimotor schemata can be considered more relevant for interaction in comparison to, for example, traditional CLIs. For example, typing on a keyboard can be considered a mere physical expression of System 2 cognition, with negligible sensorimotor schemata for typing letters and "bursts," [CNM83, p. 167] whereas exploration of an IVE using natural walking is mostly based on the parallel execution of sensorimotor schemata of varying complexity. The emerging cognitive structure based on facts, concepts, and sensorimotor schemata is enacted by users when they engage with a computer system and 'mirrors' the hierarchical composition created by the interaction technique with tasks, sub-tasks, and specified technique components [BH99] on the individual's level as part of the mental model they construct from the experienced system.

In this thesis, following the presented ideas about embodied skills and schemata, an interaction schema, as a unit of analysis of embodied interaction, is defined as:

> **Working definition: Interaction Schemata**
> Interaction schemata are flexible, hierarchical cognitive structures that are constructed by users and associated with other schemata to represent embodied knowledge about interaction with computer systems at different levels of abstraction, which enable the subconscious execution of interaction-specific procedures.

5.1.1 Basic Assumptions

Enactivism is often concerned with the question of where life, cognition, and mind begin. This is a much more fundamental level and describes a much higher dimensionality than what is useful for describing interactions in virtual environments for human users. In the following, a high-level interpretation of the enactive theory of sensorimotor contingencies is presented that incorporates the fundamental ideas within an HCI-focused framework to find a balance between the fundamental granularity expressed in enactivism and expressiveness in the context of HCI and VR (see Fig. 5.2).

To overcome the *cognitive gap* [FDP11], the expressed concern that enactivism primarily focuses on describing processes at the level of single-cell organisms, so concepts may not apply in complex human interaction, Froese and Di Paolo investigate social cognition as a complex system from an enactivist perspective. According to their framework, *social cognition*, which is specifically employed in a social context, requires a process of participatory sense-making [FDP11], which describes the process of mutually coordinated sense-making in a multi-agent system in which "novel meaning is jointly achieved through a history of breakdowns and recoveries of interactive coordination" [DPBB17, p. 243]. From an enactivist perspective, cognition can be described as the potential of a user to produce novel behavior and regulated sensorimotor coupling in the environment provided that "constitutes an emergent autonomous organization in the domains of internal and relational dynamics ..." [FDP11] which is ultimately based on "a network consisting of multiple levels of interconnected, sensorimotor subnetworks" [VTR93, p. 206]. In a similar way, in this thesis, it is assumed that a 'system cognition' has to emerge, which encompasses the enactment of the ways of thinking within the HCI system as defined by the designer of

the system. This type of cognition exists on the same level as *social cognition* (see Fig. 5.1). System cognition is influenced by the user (and, to some degree, the underlying autopoietic organism), indirectly by the decisions made by the system's designer, and by the culture and its conventions in which dedicated HCI systems for specific activities are produced.

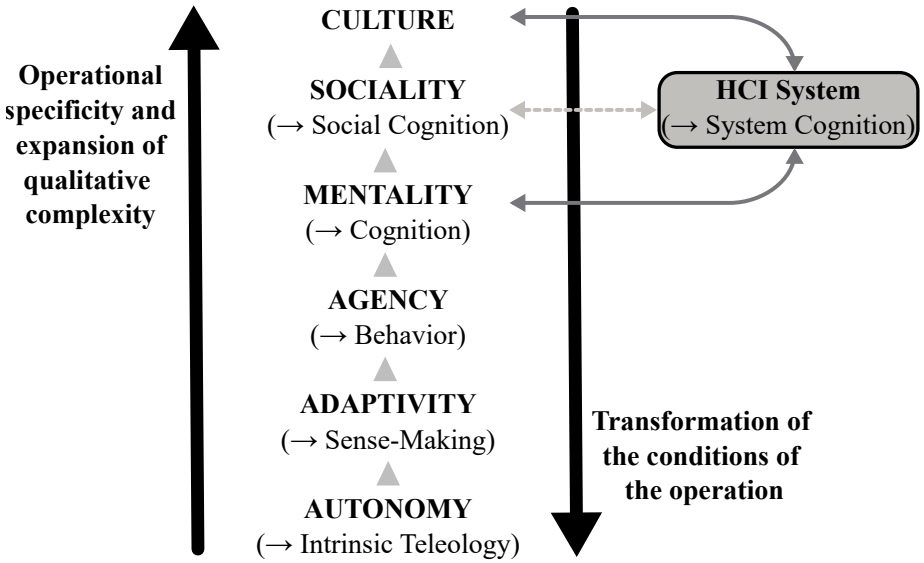


Figure 5.1: Hierarchical levels of core concepts in the enactive approach with system cognition as the subject of research in this thesis. Modification based on [FDP11].

First, the basic principles of enactivism (presented in more detail in section 2.3.3) are briefly considered for the special case of an IVE. The user and the IVE form a distinct agent-environment system that provides specific kinds of interactions. These interactions can, in some instances, be considered quite different from real-world interactions. For all tasks that involve the human body as input, the user has to enact this user-environment system and adapt to the environment to be able to skillfully cope with the challenges that are presented. On the one hand, skills can rely on abstract reasoning, cognitive operations, and symbolic input provided to the computer system. On the other hand, skills often involve procedural knowledge and a subconscious application of physical actions, which is the focus of this section. Procedural knowledge and physical skills are developed through sense-making and embodied interaction. Through sense-making, a meaning is established[24] that depends on the context of interaction and intentions of the user. This meaning typically exceeds the basic intrinsic teleology [FDP11] of an autonomous system that involves maintaining its structural integrity and fulfilling physiological needs. Often, the user is part of an activity within human culture, for example, work and play, and meaning is related to the *object* and *outcome* of such an activity system or other intrinsic reward motivations [Csi14] and subjective meanings [MH19]. In parallel to establishing a meaning, the system-specific 'system cognition' emerges that describes how a specific meaning can be achieved through a structural coupling of the embodied avatar with the virtual environment. This structural coupling can be described in terms of sensorimotor contingencies. Di Paolo, Buhrmann, and Barandiaran define four levels of sensorimotor contingencies that are briefly described in the following section [DPBB17; BDPB13].

[24] The usual term 'emerge' is not used here for reasons considered in section 5.1.5.

5.1.2 Sensorimotor Layer

The first level of sensorimotor contingencies describes the constraints of what can be sensed and perceived by an agent. Meaningful sensations and Perceptions depend on the action-perception loop and the environment in which action occurs. All that can be experienced in such a way forms the *sensorimotor environment* of a specific agent-environment system. This environment is specific to the agent-environment system but independent from internal states and actual realized actions [DPBB17; BDPB13]. This is the most fundamental and abstract level of sensorimotor enactivism. It contains only open-loop regularities in which motor action does not depend on sensory input. In the case of users in an IVE, this corresponds to all possible sensations a human could have as a result of motor actions and subsequent perceptions. In this high-level interpretation, this can also be paraphrased as 'the set of active and passive components that could form an action-perception loop.' The *sensorimotor habitat*, the next level, describes the set of all closed-loop systems that can be realized within the *sensorimotor environment* and possibly extend over time. The habitat takes into account the state of the agent and neural activity, which result from the agent's coupling to its environment. This level can, in this interpretation, be paraphrased as 'the set of all action-perception loops that can actually be realized.'

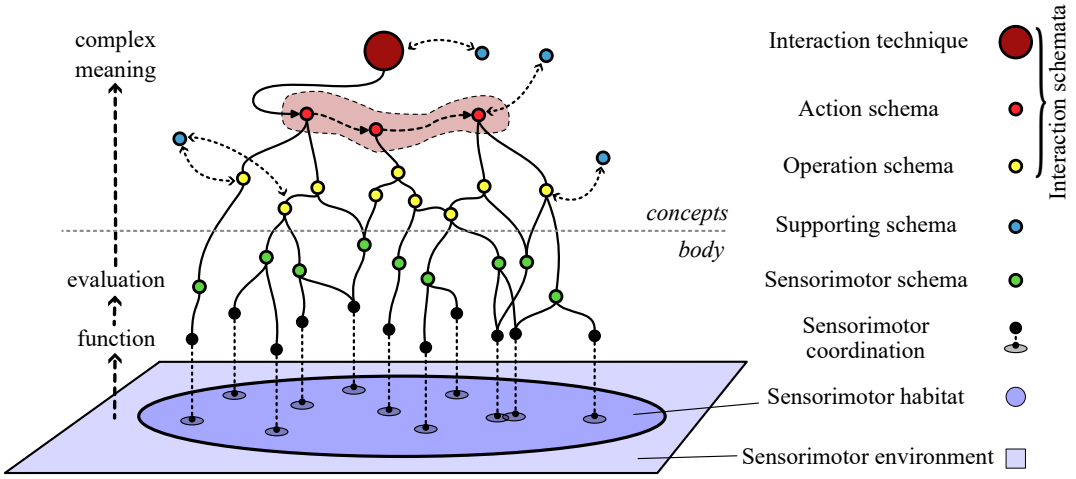


Figure 5.2: Conceptual visualization of the interpretation of the theory on sensorimotor contingencies as proposed by Di Paolo et al. [DPBB17] and the conceptual layer as a scaffolding structure for interaction in the context of HCI as proposed in this thesis.

Within this *sensorimotor habitat*, patterns of *sensorimotor coordination* are found that encompass a network of lawfully related action-perception loops that can be activated to perform a certain task or achieve a goal and functionally contribute to the activity of an agent. These *sensorimotor coordinations* correspond to all realizations of complex behavior within the sensorimotor habitat, for example, walking in different ways (slowly, fast, jumping, backward, on hands). This can be paraphrased as 'the set of all action-perception loops that can be realized and contribute to the activity of an agent.' These coordinations are normatively evaluated by the agent and hereby become "reusable, interlocking, organized sets of coordination patterns between body and environment" [DPBB17, p. 81] called *sensorimotor schemata*. The normative evaluation is based on an agent-specific and context-dependent framework that evaluates which outcomes are preferable in comparison to others. Examples of normative evaluation are "efficiency, fitness, optimality, or even subjective criteria like hedonic value" [DPBB17, p. 57]. Paraphrased, these sensorimotor schemata can be expressed as 'the set of all action-perception loops that can be realized and contribute to the activity of an agent in a (situation-specific) preferable way.'

5.1.3 Conceptional Layer

Leaving the body-centric sensorimotor framework proposed by Di Paolo et al., an additional layer is introduced in this thesis to account for interaction techniques in IVEs that are based on complex hierarchical structures and concepts. The enactive framework and the concept of sensorimotor contingencies yield the foundation for these considerations, but the complexity of interacting with 3D user interfaces in IVEs differs by orders of magnitudes to chemotaxis shown by single-cell organisms, which makes more complex structures necessary to describe full interactions. However, in the interpretation presented in this thesis, both simple movement and complex activity-focused interaction loops between humans and computers can be seen as grounded in enacted sensorimotor schemata that are coupled to an agent-environment-specific meaning and do not necessarily involve symbolic input as the main consideration for interaction and cognition.

In this thesis, the conceptual level is assumed to contain interconnected and hierarchically structured interaction schemata, which encompass i) the 'interaction technique' as the type of schema with the highest complexity (corresponding to the *method* in the GOMS model [CNM83]), ii) 'action schemata', that describe single steps in the interaction technique (roughly corresponding to the *subtasks* in [BH97]), iii) 'operation schemata,' that describe single operations (corresponding to the *operation* in the GOMS model, and the *technique component* in [BH97], respectively). Furthermore, 'supporting schemata' have an important role in associating procedural, conceptual, and declarative knowledge with other interaction schemata. An operation schema is associated with other operation schemata or, at the lowest level, with sensorimotor schemata. On this lowest level, as an example, the operation schema PRESS-A-BUTTON[25] as input action is associated with the sensorimotor schema describing the according physical movement of the finger (which is functional within the activity and has been normatively evaluated by the user to be the most preferable way of pressing this specific button). Often, such a low-level schema does not carry much meaning on its own and is incorporated into the structure of multiple schemata of higher complexity, making it 'generic' (see Fig. 5.3).

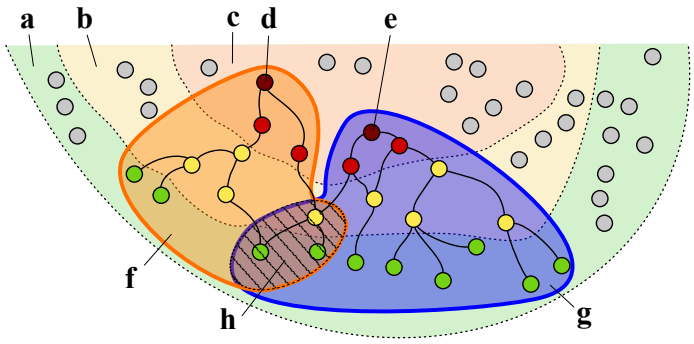


Figure 5.3: Conceptual visualization of shared low-level schemata of two interaction techniques (down to the level of sensorimotor schemata). **a**: set of all sensorimotor schemata, **b**: set of all operational schemata, **c**: set of all actions and interaction techniques, **d**: first interaction technique, **e**: second interaction technique, **f**: complete schema for *d*, **g**: complete schema for *e*, **h**: area of shared low-level schemata, for example, pressing a generic button with different functions for different high-level schemata.

Multiple operations can be conducted in parallel or sequentially to form an interaction schema with a higher degree of complexity (see Fig. 5.3). This can, for example, describe the interaction PRESS-A-BUTTON WHILE MOVING-A-POINTING DEVICE, mapping to the three low-level schemata PRESS-A-BUTTON, WHILE, and MOVING-A-POINTING-DEVICE.

[25] The small capitals letter style is chosen in analogy to the notation encountered in image schema theory.

In combination with more operation schemata, this can, for example, form the complex interaction schema '*press a button while moving a pointing device while the visual cues of a ray emerging from the pointing device are evaluated*,' which can be considered an important generic interaction schema in selection tasks and teleportation-based locomotion techniques (for example, [Boz+16]). At the highest level of complexity, the complete interaction technique is described, corresponding to a user- and system-specific meaning and incorporating all associated lower-level schemata in a hierarchical interconnected structure. At the second-highest level, schemata (here referred to as 'action schemata') can encompass sequential and parallel steps that describe a temporal and procedural multi-step sequence of an interaction technique. In this example, this corresponds to a single step within the interaction technique, such as the selection of an object for further interaction or confirming the location for a teleportation technique.

In the conceptual layer, interaction schemata are associated with supporting schemata that contain different types of knowledge about interaction. The group of image schemata forms one subgroup of supporting schemata. The application of basic image schemata as interaction concepts in HCI has been researched (see, for example, [Hur17; Hur11]) and can be applied to describe direct manipulation in GUIs, for example, moving (PATH) a file (OBJECT) into the bin (CONTAINER) [Hur11]. Not only can direct manipulation be described using image schemata (e.g., MOVING a slider OBJECT along a PATH), but complex abstract concepts can also be expressed. In IVEs, for example, the action of pressing a button can be abstracted as an image schema: PRESS-A-BUTTON. This image schema can be associated with different 'aspect expressions' [Nar97], such as START and STOP, as well as derived concepts, such as '*initiate an interaction technique*' or '*cancel an operation.*' Vice versa, these concepts are associated with different implementations of PRESS-A-BUTTON, involving a variety of actions, such as '*to initiate an action, press the trigger with the index finger*,' '*to initiate an action, press the touchpad with the thumb*,' and '*to initiate an action, press a key on a keyboard with the finger currently closest to the keyboard.*' Supporting schemata can describe declarative knowledge about the system, such as '*riding a bike is a possible means of locomotion.*' Conceptual knowledge can often be verbalized easily and corresponds to facts that the user knows or believes. Knowledge in the form of orientational, ontological, structural, and metonymical metaphors [BBN02] that transfer knowledge about one system (including reality as an agent-environment system) to another can form a supporting schema that guides interaction and facilitates the acquisition of skills. Supporting schemata can also incorporate knowledge about components of the interaction that are external to the user's body. In such a case, they describe the intentionality regarding objects (for example, 'a ball that can be thrown') and the environment ('I can open doors to move to another room'). Especially in the case of technological artifacts, the *technological intentionality* [Ver01] often enables novel forms of interaction and changes our perception of the world [Ihd90]. Object-related intentionality can be invoked by merely perceiving an object, in the words of Eleanor Rosch:

> "And given an actor with the motor programs for sitting, it is a fact of the perceived world that objects with the perceptual attributes of chairs are more likely to have functional sit-on-able-ness than objects with the appearance of cats." [RL+78]

In a similar way to sensorimotor contingencies [Sla09], which describe possible physical actions in an IVE, it is possible to think of interaction schemata as "conceptual contingencies" in IVEs that contain the conceptual ideas for interaction within a specific environment. These conceptual schemata can be adapted from the real world, or they can exclusively apply to only one virtual environment. For example, natural walking can be implemented as a generic way of locomotion in virtual environments, whereas flying and the concept

of flying-based locomotion can be considered a special interaction that is not present as a default interaction in most IVEs. One way to understand the relationship between sensorimotor contingencies and interaction schemata is through nested affordances [WS20]. At the base level, without a specific meaning, sensorimotor contingencies define how low-level actions are carried out and how stimuli are perceived (e.g., something lights up or moves). Image schemata, on the other hand, represent a higher-level set of abstracted meaningful possibilities for interaction that may include multiple sensorimotor contingencies and other interaction schemata, and they also depend on the mental model, which further contains algorithms, facts about the system, and strategies for solving complex problems.

5.1.4 Schema Coupling

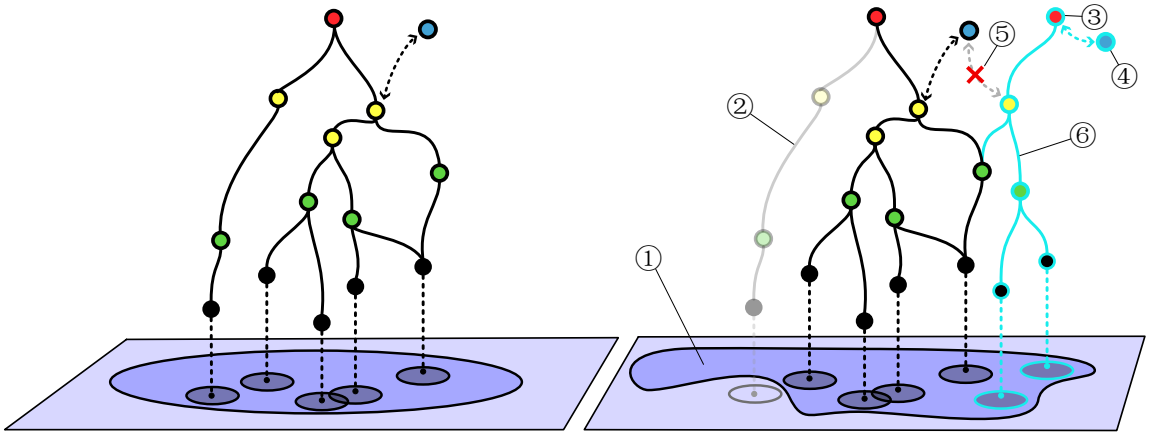


Figure 5.4: Natural interaction is typically based on the natural human sensorimotor habitat acquired in the real world (left). Novel interaction techniques can change the sensorimotor habitat (right). A super-natural interaction technique component (cyan) can be based on sensorimotor schemata that are provided by the altered habitat.

To enable users to interact with a system, the corresponding interaction schemata and related sensorimotor contingencies have to be coupled to system-specific meanings. According to Di Paolo, a sensorimotor contingency is an evaluated closed action-perception loop that corresponds to a function [DPBB17]. Super-natural interaction techniques often introduce novel functions, such as flying or shooting a virtual ray to indicate the location for a teleportation, that are coupled to certain movements of the physical body. As such, they form new sensorimotor contingencies that are coupled to a specific function, and they are only valid within the specific environment. This process can be referred to as 'coupling.' In the context of embodied interaction, Dourish describes coupling as a key component in the interaction with the world that is the responsibility of the individual user, whereas the designer, by creating the artifact, is only able to suggest and communicate ways of interaction to support the user's actual coupling [Dou04]. Coupling is not a straightforward translation of information in which "a user's immediate concerns [are mapped] onto the appropriate level of technical description," [Dou04, p.142], but instead a complex process of selecting, modifying, and evaluating action and effects of acting during interaction with an environment [Dou04].

It can be assumed that the sensorimotor environment for the human agent is static, as humans only possess a finite set of possible sensory perceptions and a finite set of possible motor actions. New super-natural interaction schemata form new closed-loop systems, which, ultimately, correspond to changes in the sensorimotor habitat (which contains all closed action-perception loops). For example, being able to control a remote pointer with hand movements forms a loop between the causal motor actions of hand movements and the

perception of a visually represented remote pointer. This closed loop can be coupled to a system function to form a sensorimotor coordination. This coordination can be normatively evaluated by the user, for example, regarding the ergonomics of the hand and arm posture, to form a sensorimotor schema. This single sensorimotor schema can be combined with other schemata to form complex compound schemata. The process of enacting the IVE-specific interaction can be related to previously enacted interactions that are grounded in the real world. The sensorimotor coupling and normative evaluation of actions and perceptions produce sensorimotor schemata [DPBB17], which are utilized for subsequent interactions.

These changes can be conceptually visualized in Figure 5.4 in analogy to Figure 5.2. New closed-loop systems expand the sensorimotor habitat, whereas missing closed-loop systems, for example, those that are expected in reality but not implemented in an IVE, reduce the extent (①). A reduction of the habitat deactivates the depending schemata (②), for example, touching virtual walls without haptic feedback. The expansion of the habitat allows for new schemata to be constructed by coordination and normative evaluation (③). In combination with other schemata, certainly not limited to the super-natural type, complex interactions such as teleportation, telekinesis, and x-ray vision can be achieved. Often, interaction is related to supporting schemata (④) that support the meaning of the interaction technique, e.g., the concept of a flying carpet conveying what can be achieved through a specific action. Other supporting schemata, especially those that are present in interactions with the physical world, can be contradicted (⑤) in VR. In the case of a flying carpet, the supporting schema 'gravity' cannot be applied, and the physical actions used to control a flying carpet, for example, tilting the upper body to control the direction of flight, can construct novel sensorimotor schemata rooted in the new technology-dependent sensorimotor habitat (⑥).

	Mechanism	Meaning	Schema	Description
I	schema integration	identical	identical	A full implementation of a schema in VR.
II	schema modification	identical	modified	Components of a schema are modified.
II a	schema reduction	identical	reduced	Only components of a schema are utilized.
II b	schema adaptation	identical	related	A different (meaning-related) schema is coupled.
III	schema coupling	identical	missing	A new schema has to be coupled to a meaning.
IV	meaning altering	related	identical	A schema invokes a different (related) meaning.
VI	sense-making	missing	missing	A new meaning is coupled to a new schema.
V	meaning absence	missing	identical	A schema-meaning coupling is not implemented.

Table 5.1: Description of eight identified schema-meaning coupling mechanisms for interaction in VR.

For VR, we propose distinct mechanisms for coupling that can be described systematically in terms of the existence of both meaning and schemata prior to interaction in the IVE (see Table 5.1). 'Meaning,' in this context, is used to describe that an action has significance to the acting agent who is working towards a goal or outcome in an activity. It is an identifiable and describable intention that is related to a discrete or abstract challenge. This meaning is coupled with multiple instances of interaction schemata that form, in their particular complete structure, a distinct way of thinking about a problem, achieving some goal, or satisfying a need. In the case of locomotion as meaning, for example, possible interaction schemata could be 'natural walking,' 'teleportation,' or 'flying.' These schemata are further tunable (see 'tuning' in [Rum17]) to enable application in diverse situations, and they can be modified (see 'restructuring' in [Rum17]) if an application is not directly possible, which produces new ways of acting. To acquire the IVE-dependent skills in an IVE, the user has to attune to the properties of the environment. The level of difficulty

of attuning depends on the incorporated coupling mechanisms of an interaction technique and intrinsic factors related to the complexity of sense-making (see Fig. 5.5). Typically, schema integration allows for a direct application of acquired skills without sense-making and can be considered, therefore, the simplest form of coupling. Variations, such as schema reduction, modification, and altering, require some sense-making depending on the degree of variation. A missing schema or a missing meaning always requires sense-making.

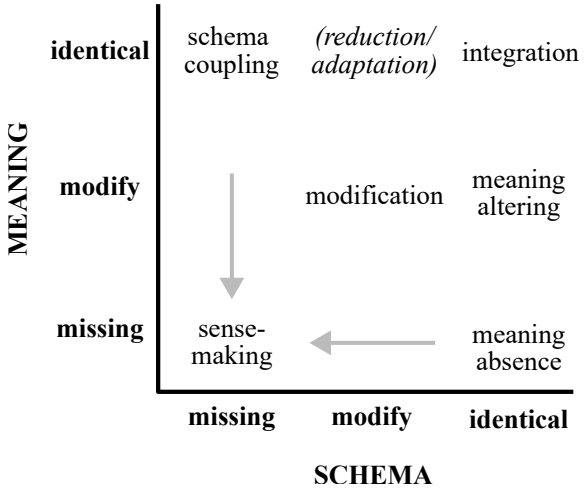


Figure 5.5: Schema coupling mechanisms in coordinate system.

A direct manipulation of objects is a meaningful real-world task that is implemented in many VR applications and relies on a combination of coupling mechanisms. In the real world, basic object manipulation is achieved by moving a hand to the spatial location of an object, applying skillful prehensile gestures that are aligned to the shape of the object [Fei+15], and transforming the hand while maintaining and adjusting the prehensile gesture. These actions are perceptually guided through vision (hand-eye coordination) and also haptics (sensation of the object in the hand). Changing the location of an object is coupled to a meaning, for example, skillfully coping with the environment by inserting a key into a lock to open a door. This meaning can be adapted in an IVE by implementing a virtual key that opens a virtual door (*schema integration*). Different mechanisms and the involved skills and coupled sensorimotor schemata that are applied to implement this task can be described (see Table 5.1). By replacing the prehensile gesture with the press of a button, the involved sensorimotor contingencies are simplified and reduced from the movement of multiple fingers to a simple abstract action (*schema reduction*). While the haptic feedback is missing (*meaning absence*), new visual signals can be implemented, such as highlighting the selected object (*sense-making*). The transformation of the hand as the primary means of transforming the object is maintained, and the hand-eye coordination provides feedback on actions within the environment (*schema integration*).

But IVEs do not only allow for the recreation of realistic interactions. A very successful supernatural locomotion technique is point-&-click teleportation [Boz+16]. The meaning of locomotion is grounded in our everyday interaction and experience of the real world, however, there is no inherent sensorimotor coupling for teleportation, so a new sensorimotor schema has to be coupled (*sense-making*). Often, the task of teleportation is initiated with a pointing phase, in which a virtual laser beam or similar visualization may be displayed to facilitate the indication of a location. Although pointing with fingers is a typical real-life task, extending the human finger or hand with a visual element, which supports the initial meaning, is not (*meaning altering*). Modifying the effect of a sensorimotor schema can also be observed as scaling of walking speed to increase speed [IRA07], or extending the reach of one's arm for remote object manipulation [Pou+96a] (*meaning altering*). IVEs can

also implement schemata of actions that resemble certain aspects of the original interaction. Real-world walking can, for example, be approximated with leaning-based techniques [Har+14], walking-by-cycling [Fre+20], and finger walking [Kim+08] (*schema adaptation*). IVEs also enable completely new meanings that have no direct counterpart in the real world, for example, creating or removing virtual objects or changing the scale and color of objects (*sense-making*). Developers have great freedom in designing these interactions, as they cannot be related to the real world. In many cases, GUI-inspired virtual menus are utilized to perform such actions, which only require the user to learn a new schema to open a menu (*sense-making*), often implemented as a hand gesture or controller button, and a way of selecting menu items, for example projected virtual pointers (*sense-making*) or direct touch (*schema integration*[26]).

5.1.5 Internalizability

The acquisition of new interaction schemata and integration in the individual's cognitive structures can be described as 'internalization.' Internalization, as used in this thesis, is related to how the term 'internalization' was used by Piaget to describe the construction of knowledge during cognitive development in the form of integrating new schemata into the existing knowledge. Di Paolo et al. [DPBB17] emphasize that Piaget's theories, on a fundamental level, resonate well with sensorimotor enactivism; however, in enactivism, the abstract reasoning in adults is not seen as the final stage or goal of cognitive development. In this thesis, it is assumed that, in a similar way as cognitive structures have to develop to account for real-world interaction, they also have to adapt to account for specific user interfaces and their employed interaction techniques. To successfully achieve system cognition, as it has been called in the previous section, users have to internalize the schemata that are implemented in the system.

The term 'internalization' is also encountered in activity theory and social science to describe a similar process that is, in contrast to Piaget's idea of individual internalization, more focused on the social component of acting within a community. In activity theory, knowledge is shared between subjects within a community through a process called the 'zone of proximal development,' and practices produce externalized knowledge that can be internalized by an individual. For example, a factory worker directly internalizes how to control a specific machine by observing other workers controlling it. While the social component is not necessarily present when a user learns to control a computer system, this idea still points to an HCI-specific fact: The environment and interaction techniques are designed by another cognitive agent and, therefore, based on their internal cognitive structures. When a VR designer (or developer/researcher) implements a specific interaction, this interaction is aimed at an intentional aspect of the system interaction that corresponds to a distinct meaning. This meaning has been (what could be called) 'pro-actively enacted' by the designer, which means that the designer did not only take an active role in the sense-making, and subsequently, a meaning emerged, but instead, meaning preceded the sense-making, and the designer had to identify and specify ways to allow a user to enact this specific meaning in the intended ways. Enacting this developer-dependent meaning, including the procedural ways of achieving this meaning, can be called 'internalization', which accounts for both the construction of cognitive structures (in analogy to the use of 'internalization' in Piagetian constructivism) and the indirect transfer of cognitive structures from the designer's conceptual model of a system to the user's mental model of a system (in analogy to the use of 'internalization' in activity theory).

Some interfaces externalize information about interfaces. For example, crib-sheet [KMB94]

[26] Direct touch of buttons can be considered a real-world interaction considering the wide-spread use of touch displays outside of IVEs nowadays.

and octopocus [BM08] guide users through gestural interaction by providing additional external visual cues. Furthermore, tooltips as well as manuals can be seen as externalized knowledge about the interaction with a computer system, which are intended to instruct users on how the system is used. However, a complete internalization that does not require any external resources can also be achieved in many cases. In such a case, the interaction schema is *discovered* [ACD09]. The idea of internalization has been expressed by Bibby and Payne in their *internalization hypothesis*:

> "We suggest that [informational and computational equivalence][27] can be extended from external representations to mental representations held in long-term memory and, thus, help understand the functional distinctions in declarative knowledge acquired from different instructions. As external forms, different instructions have both informational and computational properties affecting their use. We suggest that memorizing the instructions will preserve these informational and computational properties. The resulting declarative knowledge will lead to patterns of search, recognition, and inference that mirror exactly the processes that operate on the external forms. We refer to this conjecture as the "internalization hypothesis."" [BP93]

In this formulation, Bibby and Payne explicitly focus on declarative knowledge. In this thesis, it is assumed that this hypothesis is true not only for declarative knowledge but also for procedural knowledge and embodied cognition in 3D user interfaces. This encompasses procedural knowledge that can be transferred to internal structures through different means, such as action-oriented minimal representations and pre-intentional acts, as well as more complex multi-step physical interactions.

Furthermore, the degree of internalization a user has achieved for a specific schema can vary. Typically, skills need to be rehearsed over an extended period of time to become completely automated and subconscious [CG17; Swe03]. Corresponding to the idea of overlearning [Nel+82], skills can be considered to be, at some point, 'fully internalized,' when a further improvement of recalling and applying skills cannot be achieved. In such a case, the user reaches, in enactivist terms, *attunement* to the system or a specific task, which is characterized by mastery of the involved skills. Considering the interaction with artifacts, fully internalized schemata correspond to the idea of *embodied* and *background* technology relations presented in post-phenomenology [Ihd90; RV15]. For example, driving a car reaches, for experienced drivers, a technological transparency and readiness-to-hand[28] in which they are, phenomenologically, "driving down the road, not operating controls" [WF86, p. 164]. *Alterity* technology and *hermeneutic* technology can also be seen as supporting schemata that guide interaction, for example, how and which problems can typically be engaged. Fully internalized interaction encompasses the activation of a context-, user- and environment-specific virtual field of interconnected interaction and support schemata that provide the basis for cognitive processes.

According to Sweller, processing complex information is only possible by constructing corresponding schemata that encode perceptions and actions efficiently [Swe03]. Complex interaction with technology can often be considered a domain that relies on fully internalized interaction schemata as fundamental interaction components required for interacting with complex information. For example, typing on a keyboard or using a mouse as a pointing device in a desktop computer system can be considered a fully embodied and transparent

[27] Referencing an analysis regarding diagram visualizations by Larkin and Simon (see [LS87]).

[28] This is a Heideggerian term in phenomenology (original German: '*Zuhandenheit*'), referencing the perception of an object based on one's intentionality and the intended practical use. The opposite term, 'present-at-hand' (original German: '*Vorhandenheit*') describes the perception of an object that is perceived free from intentionality.

interaction for expert users, and controlling a system using these means does not require additional effort, which enables users to fully concentrate on the information relevant to current tasks. VR systems that include the user's body and hand can be considered more corresponding to real-world experience, which allows for a much more enactive approach to HCI. Instead of typing a command and physical actions as a mere expression of cognitive processes, users can physically manipulate objects, which largely corresponds to the activation of schemata acquired in the real world. Recreating realistic interaction, or providing means of interaction that allow enaction via comprehensible action-perception loops for novel interaction, supports the acquisition, modification, and application of schemata. Missing support of schemata-based interaction creates "just one more domain in which a skilled agent may act and perceive. ... The modes of sensing and interaction supported by current technologies often remain limited and clumsy, and this turns the user experience into that of a kind of alert game player rather than that of an agent genuinely located inside the virtual world" [Cla08]. A sophisticated interaction design that shows a high internalizability can reduce this 'clumsiness' and help users to feel as if they are actually integrated into an IVE and embodying their avatar instead of controlling an application from the outside.

The internalization of different interaction techniques can vary, and the ability to fully internalize the controls of an interaction technique is, in this thesis, called 'internalizability.' Internalization requires the construction or adaptation of cognitive structures to account for experiences, and, depending on the design of interaction, this can be more or less difficult. Rumelhart describes several ways of constructing and modifying schemata [Rum17] that can also be utilized to describe procedural knowledge. According to him, in a process called *accretion*, past interactions and their *fragmentary memory traces* can be transformed into schemata. Schemata can also be modified by *tuning*, either by transforming constant aspects into variables and vice versa, or by adapting variables to specific conditions. In the*restructuring* process, either through *pattern generation* or through *schema induction*, cognitive structures are constructed from experience to enable the construction of novel structures. These means of internalization can be considered the underlying process of learning a new interface interaction. In this thesis, internalizability refers to a subtype of learnability that specifically encompasses the following properties:

- *Flexible patterns*: The inclusion of schematic knowledge consisting of constants and variables that are modified to account for diverse factors in non-deterministic environments. They primarily encompass concepts of action that can be tuned to situations rather than discrete facts.
- *Automation*: Internalized schemata affect behavior and system cognition in a subconscious and automated way and are intended to be not consciously retrieved during interaction.
- *Techno-social constructs*: Interaction schemata have been intentionally designed by another person for a specific context or task, which users integrate into their own cognitive structure. Users can either be instructed about the procedural skills involved in the interaction (in the form of externalized knowledge) or discover these themselves.
- *Action-oriented*: They contain knowledge about meaningful acting within an environment, including conceptual and embodied knowledge that is incorporated into the user's body schema.
- *Association*: Schemata form a network of associated supporting schemata, other operational schemata, and sensorimotor contingencies (involving perception, action, or both).

To summarize, internalizability can be defined as:

> **Working definition: Internalizability**
> Internalizability describes how easily and effectively a user can fully incorporate the interaction schemata representing a specific interaction technique into their cognitive structures.

5.2 Schemata-Based Interaction Cycle

The assumptions of information processing theory can be considered the traditional framework for cognition in HCI, in the form of analyzing the transmission and manipulation of mental representations. Different models based on this framework describe the interaction cycle between a user and a computer system in the form of System 1 cognition. Enactivism, on the other hand, and the presented concept of interaction schemata diverge from this assumption and lay their focus on interaction cycles based on System 2 cognition. Strack and Deutsch propose nine theses that describe the dual process theory of cognition in the context of social interaction [SD04]. Their theses can be summarized (the numbering and naming are directly adopted from [SD04]) as:

1. *Basic Assumption*: Behavior is produced by two distinct systems: an impulsive system ('System 1') and a reflective system ('System 2').
2. *Parallel Operation*: The reflective and impulsive systems operate in parallel, and the reflective system only engages when needed.
3. *Capacity*: The impulsive system requires low cognitive resources, whereas the reflective system has high demands.
4. *Relations between elements*: Cognitive elements are connected either by associations (System 1) or by semantic relations (System 2).
5. *Final Common Pathway to Behavior*: Both systems are capable of activating behavior in the form of *sensory-motor clusters* on the impulsive level.
6. *Precursors of Behavior*: Behavior is produced through the spread of activation of schemata (System 1) or by rational decision-making (System 2)
7. *Intending*: The reflective system monitors the impulsive system in a dedicated process for conditions that allow for the implementation of behavior. This reduces the required cognitive capacity.
8. *Motivational Orientation*: Positive and negative effects orient the impulsive system toward either *approach* or *avoidance*.
9. *Compatibility*: When information processing, experiencing affect, and the execution of behavior correspond to the motivational orientation, they are facilitated.

The underlying mechanisms of System 2 cognition cannot be directly observed. However, based on the literature on cognitive science and HCI, certain steps in System 2 cognition can be described. These are, however, not interpreted as discrete modules in a cognitive architecture, as typically assumed in information processing theory. From an enactivist perspective, the steps described in the following section are merely used to structure certain aspects of the spread of neural activity across complex sensorimotor networks and virtual fields that are activated in response to observations in the environment in an automated and unconscious ongoing process.

5.2.1 System Cognition

The user in an HCI system is, from an enactivist perspective, a biological organism with individual needs and meaning in the world, which forms the basis of interaction. Meaning and needs can be physiological, but often, they are more complex and depend on the activity of the user or personal factors, such as cognition, conation, and emotion [Wei+12], forming a "mentality beyond metabolism" [FDP11]. From these perceived needs and meanings, an initial intention for action is formed, often based on internalized schemata for interaction that are coupled to agent-specific meanings that have been enacted in previous engagements with this or similar systems. The knowledge for interaction can be top-down if conceptual ideas are primarily used to guide the interaction. When a user has experienced a contemporary IVE before, they suspect that natural walking will be a typical means to explore the virtual environment. Interaction can also follow a bottom-up approach if performing arbitrary physical actions leads to the exploration of meaningful interaction schemata. For example, some first-time VR users may be unaware that freely moving their heads is a typical way to change the 3D perspective in a virtual environment, as this is typically not possible in other types of media, such as traditional 2D screens.

These aspects contribute to system cognition, which describes how a user acts and thinks within a defined system. It encompasses both declarative and procedural knowledge that can only be applied within the context of a specific technical system. This knowledge can range from abstract information, such as facts and algorithms, to schematic knowledge related to interaction in the form of interaction schemata. The way users act and think within such systems typically differs significantly from basic, non-technological interactions, and their motivations and intentions are not necessarily tied to real-world needs. Systems often combine abstract information and interaction schemata to enable users to solve tasks. For instance, in traditional computer systems, the input of symbols can be broken down into abstract components processed in System 2 cognition and interaction schemata in System 1 cognition that enable the symbolic input. In contrast, VR systems can describe entire interactions through interaction schemata; for example, locomotion is achieved through natural walking, and object manipulation is facilitated through direct manipulation.

If a completely new agent-environment system is engaged, the corresponding system cognition has to be constructed. In many cases, other instances of system cognition that users perceive as similar can be activated as a first approach to understanding a system based on prior knowledge. This includes ways of interaction in the form of metaphors and conventions, and, furthermore, motivation for interaction and emotional responses to observation. Driving a virtual car, for example, is not only about transferring controls from reality to the IVE but also involves a (post-)phenomenological perception of the environment, such as how distances are perceived and a 'mode' of engaging with the world. During cognition, new ways of interaction can be acquired that modify the system cognition, either by internalizing the implemented schemata or by learning strategies for performing information processing for interaction knowledge that cannot be easily stored as schemata. System cognition is inherently connected to aspects of the emerging pattern of self, such as cognitive aspects, embodiment, and agency. In this enactivist modeling of HCI, it is important to emphasize that a user is not merely an abstract symbol-processing entity. The user remains a living organism in a lived body with a subjective experience of existing in an environment, which leads to emotional responses, meaningful impressions, and individual intentionality. The overall experience of interaction with a certain pattern of self within a specific context leads to a phenomenal experience of the self in a temporary agent-environment system constituted by the technological environment.

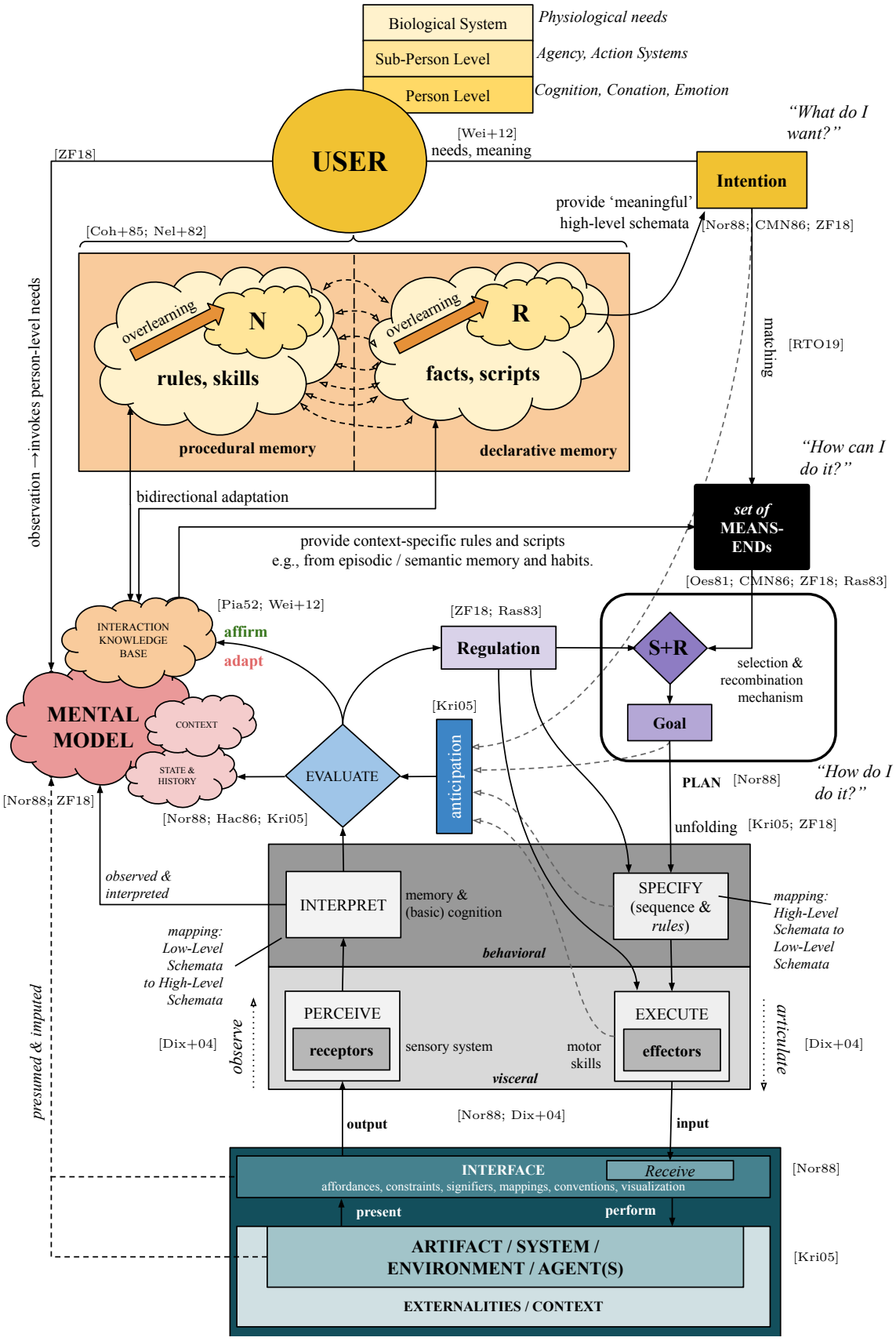


Figure 5.6: The combination of different HCI models allows for a detailed analysis of aspects of schemata-based HCI.

5.2.2 Mental Model

An essential component of system cognition is the storage and retrieval of metaphors. Metaphors include image schemas (see section 2.3.1), such as directions, forces, or 3D positioning, as well as declarative knowledge, such as the intended purpose of certain objects in the real world (*structural metaphors*) or concepts that are specific to the IVE. Interaction techniques can be broken down into reusable components that may reappear as conventions across different applications. Some of these components are assigned application-specific functions, while others contain standardized methods for completing tasks. As a result, system cognition encompasses both declarative elements that describe what is possible within a system and procedural knowledge on how goals can be achieved. Users can access this knowledge during their initial use of the systems to guide their basic interactions as part of the mental model.

The mental model represents various properties of a system and provides concepts and schemata for interaction [Ras87; CO88; Ber+08], allowing for both reflective action planning and impulsive schema application. Especially in the case of familiar actions, less demanding skill-based behavior [Ras87] and action-selection [LJ+17] are employed, which consist of recalling and applying schemata for interaction instead of utilizing more resource-intensive strategies for decision-making. System 2 is only activated in cases in which a direct mapping of perception and action using System 1 is not possible, and attention needs to be shifted towards the selection and recombination of schemata. Categories in which employing System 2 is required are, for example, planning and decision-making, troubleshooting, ill-learned or novel action sequences, dangerous or difficult situations (as long as not time-critical), and the overcoming of habits [NS86]. Schemata, both those that are believed and those that have been verified in the course of many interaction loops, form an adaptive *interaction knowledge base* [Ben93] as part of the mental model that accounts for the current configuration of the agent-environment system and specific possible interactions. Interaction knowledge is formed through previous interactions within a similar context. It encompasses the virtual field of interaction schemata in a given context that may lead to a specific outcome. When a user is confronted with an IVE for the first time, the mental model adapts scripts and interaction schemata from previously experienced similar applications as 'weak' candidates of what may be possible in an unfamiliar IVE. The system can also be designed to display a system-specific *interface* that provides cues and restrictions on what can be *performed* [Nor88]. By applying different scripts and schemata, and subsequently observing the outcome, the user builds and constantly refines the interaction knowledge base for a specific IVE to contain 'strong' candidates for interaction. Furthermore, the interaction knowledge base can be interpreted as sedimentation that enables inferencing schemata for interaction by forming abductive hypotheses [Wei+12], which can also be applied in other, similar IVEs.

Virtual actions in the virtual field are based on complex interconnected networks of interaction schemata that include procedural knowledge as sensorimotor and interaction schemata (rules, skills) and declarative knowledge (facts, scripts) in the form of support schemata. Munro et al. list *structural*, *procedural*, *and behavioral knowledge* [Mun+02] that can be considered support schemata for interaction. In their description, *structural knowledge* contains information about relationships between objects, such as *part-whole knowledge* (multiple simple objects as parts of a more complex object), *support-dependent knowledge* (the physical structure of objects that is affected by gravity), and *containment knowledge* (the spatial location of an object within another). *Procedural knowledge*, on the other hand, describes how an action has to be carried out, such as *task-prerequisite knowledge* (preparation of an initial state for action), *goal-hierarchy knowledge* (hierarchical structure of goals and sub-goals for an action), and *action-sequence knowledge* (the right order of

sequential steps in an action). Last, *behavioral knowledge* contains knowledge about the general interaction with objects, such as *cause-and-effect knowledge* (anticipated system responses for specific actions), *function knowledge* (the goal-directed use of an object), and *systemic-behavior knowledge* (generalization of knowledge). Both types of knowledge can be overlearned to form a solid scaffolding for interaction, which has a positive effect on future application and automation of skills [Nel+82], either as fully internalized sensorimotor schemata (e.g., grabbing a cup to drink), or as firmly believed facts (e.g., 'objects fall down due to gravity'). In their fundamental structure, interaction schemata and supporting schemata are constructed from associated primitive functions, such as *p-primes* [Der96] and *sensory-motor actions* [Fis80], to represent complex actions and experiences in the form of *image schemata* [Lak12].

For traditional computer systems, Don Norman criticizes the idea of GUI-controlled powerful personal computers as a multi-purpose device and suggests: "rather than trying to make a complex machine easy, the better way would be to make a simple machine in the first place" [Nor98, p. 73]. This statement aligns well with the concept of grounding systems and interaction techniques on simple schemata that are well-suited to solve specific tasks, rather than general systems that require more cognitive effort and symbolic processing to achieve an intended outcome. In many VR applications, a multi-purpose input device with reconfigurable buttons used as the primary input device for interacting with the virtual environment can be interpreted as such a complex machine. It enables interaction on a symbolic level by utilizing simple sensorimotor schemata such as coordinated hand-eye movements for pointing, tilting, and pressing buttons. But to enable handling this complex device, VR applications are often required to describe the function of certain buttons or highlight which button has to be pressed to achieve a desired interaction. This implies relying on a too complex mental model required for interaction instead of implementing simple sensorimotor schemata as a common approach in contemporary VR interaction design.

5.2.3 Articulation

From the user's intention, a virtual field of possible actions is matched as a set of means-ends with the user as a decision-making entity that evaluates which interaction schemata are most likely to produce the intended outcome, either by selecting an adequate schema or by recombining multiple schemata into a new schema to account for situative effects. In this selection [CNM83] and recombination process, the most promising candidate for interaction is determined from a *set of means-ends* as *goal* [Nor88; CMN86] of an instrumental action [Wei+12] ("What do I do?").

Selection describes a form of lookup, how a specific goal has been achieved before, and applying scripts in the same way. In such a case, the event-based *stimulation patterns* that produce some form of *sensory code* are automatically and subconsciously translated to *motor codes* and *excitation patterns* that produce an appropriate reaction [Pri97]. The automatic selection process can be supported regarding, e.g., accuracy and speed [LJ+17] by a high stimulus-response compatibility, for example, when sensory input and motor action are spatially similar [Pri97], or stimulus and response are kept constant instead of switching tasks [LJ+17]. Norman and Shallice propose [NS86] a model to describe the selection process of schemata. In their model, the *horizontal thread* is a central element that connects the activation of schemata to *perceptual triggers* that are activated by *sensory-perceptual structures* [NS86]. According to their model, three different mechanisms influence the selection of schemata: *vertical thread influences*, *contention scheduling influences*, and *trigger condition influences* [NS86]. *Vertical thread influences* are intrapersonal factors such as motivation and attention that increase or decrease the activation of specific schemata. For

parallel processing of tasks, the activation of schemata is inhibited or excited by *contention scheduling*, which enables the cooperative use of shared schema structures and limited resources. *Trigger conditions* describe the required conditions that enable the activation of a schema, for example, in timing-critical actions. *Recombination*, on the other hand, is the generation of a sequence of possibly parallel actions that may lead to a desired result. This process is complex as it is influenced by numerous internal and external factors and can be described as an application of "meta-cognitive heuristics" [ZF18]. For example, the activation of a teleportation event is not standardized, and different methods are currently used to initiate teleportation, e.g., pressing the trigger or a specific button on a controller. If the system-specific input is known, a user only has to recall the associated sensorimotor schema of pressing the correct button (selection). If this is not possible, the user may also try out different variations of the teleportation as rule-based behavior [Ras87] until the correct button is identified. However, in both cases, the user needs to possess knowledge of the possibility of employing teleportation as locomotion in the IVE , which is, ultimately, related to the system cognition.

In System 2 cognition, the goal unfolds [Kri05] by *specifying* a sequence of hierarchical actions [ZF18], which are *executed* [Nor88] to *articulate* them as *input* to the system [Dix+04]. From a psychological perspective, the *specification* is conducted on a behavioral level and the *execution* is carried out on a visceral level [Nor88]. In System 1 cognition, on the other hand, the mapping from observations to responses is more direct. The goal of actions is known, and the sequence of actions is not consciously specified but rather subconsciously recalled. Due to the final common pathway of behavior [SD04], the articulation seems similar regardless of the employed system, and both create an anticipation for intended system reactions. *High-level schemata* are decomposed into a sequence of *low-level schemata* which can be executed in a planned order or in parallel [Oes81]. The computer system receives these *inputs* and translates the user's commands to a system-specific reaction.

The user input that controls a super-natural interaction can often be interpreted as a 'recombination' of real-world schemata and previously acquired interaction schemata that form a new schema. This reduction can be further analyzed, for example, for teleportation in IVEs. Research shows that teleportation is an efficient and pleasant way of traveling in IVEs [Boz+16; LLS18]. Even though teleportation is not possible in the real world and contradicts our real-world experience, this super-natural interaction can reliably be applied with only a short time of training. As a result, many IVEs implement this interaction technique to account for limited real space in a possibly unlimited virtual environment. To use teleportation, the user first needs to retrieve the script from the context-dependent set of means-ends that contains the information: 'In this context, teleportation can be used to change location.' In this example, this script contains a sequence of two subscripts that point to individual rules. First, the user indicates the target location using a hand-held pointing device (e.g., a controller) and initiates the teleportation action (in this example, pressing a trigger button). Pressing trigger buttons is a generic action that can be mapped to different functions. The procedural actions (rules) and sensorimotor schemata required to press a button are fairly simple, which makes this schema easily overlearnable. The indication can be supported by visual cues such as virtual rays that enhance the pointing direction and a virtual target at the selected location. By interpreting these visual cues, the spatial movement of involved parts of the body is continuously regulated until the target location (which can also be a sub-location on the way to the final destination) is indicated. The second phase is the confirmation of teleportation, in this case, by releasing the trigger. After confirmation, the user input is evaluated by the computer system, and the user's position is shifted accordingly.

5.2.4 Observation

The reaction is *presented* to the user, which may produce an output that can be sensed by the user, such as audio or visual elements, or in an indirect and subtle way without further evidence of produced effects. The user *perceives* (on the visceral level) the system's feedback and *interprets* the result (on the behavioral level) [Nor88]. Again, perceptual schemata can be applied to interpret a system response more easily. This interpretation is compared to *anticipations* [Kri05] on different levels (From highest to lowest: What did the user want initially (*intention*)? Does the selected means-end script lead to the desired outcome (*goal*)? Is the application of schemata done correctly (*specification*)? Did a specific rule produce the expected outcome in this context (*execution*)? On a lower level of evaluation, a continuous *regulation* bypasses the high-level conscious cycle of forming a goal to control the application of rules in a mostly subconscious and automated way or on a semi-conscious level [ZF18], for example on the 'sensomotoric regulation-level' and on the 'perceptive-conceptual'-level [Hac86]. On this lower level, the user expects an anticipated action-induced perception in analogy to prior experiences [KFD15]. On the higher 'intellectual' level [SFS12], the approach of reaching a goal and the employed schemata are re-evaluated and corrected, if necessary. Schemata are constantly observed regarding their goodness of fit [Rum17] and are tuned to the specific situation.

According to Krippendorff, the engagement with an artifact consists of three phases in which the artifact, in this case, an interaction technique, is observed and interpreted: [Kri05]: i) Successfully identifying what an artifact is and its context (*recognition*), ii) identifying how an artifact is used and what can be achieved (*exploration*), and iii) the natural handling of an artifact with conscious attention lying on the consequence of actions and not the handling itself (*reliance*). In this thesis, it is assumed that these phases are also present for interaction techniques in VR, and their fundamental working principle is the enacting or application of interaction schemata. Often, when first encountering an unfamiliar system, *recognition* and *exploration* can be conducted by applying schemata that are already known. Design strategies for interactive systems, such as comprehensible mapping, discoverability, signifiers, constraints, and feedback, facilitate this step. Metaphors and affordances, both of which are based on previous knowledge, enable a novice user to 'guess' which actions are possible. Often, conventional metaphors are implemented to enable users to intuitively apply generic schemata that have been enacted in previous interactions with computer systems, in VR interaction, for example, pressing buttons on controllers or moving limbs. Only when familiar schemata cannot be applied to engage with an interface, an additional cognitive System 2 process is initiated to modify schemata or construct new ways of interaction. Achieving *reliance*, on the other hand, is only possible through internalizing the required interaction schemata.

In case of an intended system reaction, the interaction knowledge base is strengthened, and rules and scripts are *affirmed*, whereas, if the system reaction deviates from the expected behavior, the *mental model* needs to be updated by *adapting* the model to evaluation [Pia52]. A deviation from expected behavior can be described as a 'disruption' that leads to a transition from the 'reliance' phase of using a system to an 'exploration' of the system-specific functionalities [Kri05]. This process depends on the ability of schemata to evaluate their goodness of fit [Rum17]. During the exploration phase, schemata can be tested and identified as potential forms of interaction that enable users to achieve their goals. In the reliance phase, primarily those schemata whose goodness of fit has been repeatedly confirmed are used and form a basis for the mental model. Continuous monitoring ensures that schemata are evaluated throughout the interaction, which allows for dynamic adjustments or abandonment if necessary.

CHAPTER 6

CLASSIFYING INTERACTION: THE ICE CUBE

6.1 Motivation

To make the conceptual considerations of the previous chapters applicable in human-computer interaction research, this section proposes a model that unifies these ideas to provide a simple yet expressive theoretical construct that can be employed in subsequent research: the ICE cube. As a result of inductive reasoning [Kuc19], it is a representation of one possible perspective on super-natural interaction techniques, which is grounded in the literature review and the enactive approach and related frameworks. The model is based on three dimensions: internalizability, congruence, and enhancement, which are conceptualized as orthogonal axes in a Cartesian coordinate system. This enables the localization of various interaction techniques within a three-dimensional space. These axes are considered continua, and their characteristics are briefly described in this section. To verify the model, a study is conducted to examine the extent to which experts position interaction techniques similarly within this 3D space of interaction design.

6.2 The ICE Cube

6.2.1 Dimensions

Internalizability The approach for definition is to interpret the property super-natural as an extension to the property 'natural,' as previously proposed by Steinicke [Ste17b] and Lubos [Lub18]. Internalizability is used as a measure of how naturally people can use an interaction technique to achieve a function. Regarding cognitive processes, Sweller wrote: "Our cognitive architecture has evolved so that very high element interactivity material encompassing large amounts of information can *only* be handled when incorporated in schemas" [Swe03, p. 224]. We believe this principle also applies to super-natural interaction.

The term "natural interaction" is a similar term to internalizability and has been used to describe intuitive interfaces [WW11]. Although the term originated in natural language interfaces [Mil75], the use of the term is still debated nowadays [Nor10; Fu+18; Sza19]. We view internalizability as a property of interfaces that ultimately leads to natural interaction, independent of using real-world metaphors as the foundation for interaction. In our modern technology-infused world, many interactions with our environment bear little similarity to interactions with nature but are internalized and part of our natural interaction behavior [JRG14]. Furthermore, not every interaction with the real world is automatically highly internalizable. Playing a musical instrument or flying an airplane, for example, require years of training to master, and these skills are, therefore, reality-based but not easily internalizable.

In the presented framework, internalizability emphasizes this distinction between easy-to-learn and easy-to-apply on one hand and interaction based on the natural environment

around us (which is part of the congruence dimension) on the other. This concept also resonates well with the notion of 'attunement' in enactivism, which describes the state of an agent that has successfully adapted to a specific environment by enacting the required skills to respond in accordance with organismic and environmental demands. Such attunement also aligns with System 1 cognition and schemata-based interaction, wherein acquired skills are applied seamlessly without conscious effort.

Based on their prior internalized knowledge, users may perceive interfaces differently in terms of their internalizability. For example, a chopstick interface [Ke+20; Kit+99] can be an adequate choice for input modality if it can be assumed that the intended users have already undergone sufficient training in the utilization of chopsticks, which is largely dependent on the cultural background. This makes the perception of the 'naturalness' of an interaction technique a highly subjective property. However, 'natural' can also be considered an objective property of an interface if most of the intended users perceive the implemented interaction techniques as 'natural' based on their already internalized interaction schemata. In this regard, the internalizability of an interface, which describes the effort required for achieving a full internalization of the designed and implemented interaction schemata, is the key element of acquiring new interaction techniques that may, after internalization, feel 'natural.' For example, following this definition, it can be proposed that *digital natives* perceive the interaction with digital content as natural because they have fully internalized the typical ways of interacting with digital content from an early age, which facilitates the acquisition of novel related digital skills and navigating the digital world.

In the literature review regarding the properties of super-natural interaction (see Section 3.2), two perspectives have been identified: natural interaction (feeling or body-related) and the inclusion of story elements. Internalization can be used as a concept to unify both the physical actions involved in performing a task and the cognitive patterns of thinking. In both cases, the internalization of new schemata for interaction benefits from pre-existing structures that are recombined to form a new interaction schema. In the case of what Lubos calls *natural human mechanisms*, such as "grabbing, walking, touching, speaking, looking" [Lub18, p. 13], these are well-enacted sensorimotor schemata that can be integrated to facilitate the acquisition of new skills by forming a scaffold for coupling physical actions to the skill-specific meaning. In the case of story elements, these can be thought of as familiar high-level schemata that represent specific meanings and implications for the relation of the user and the environment. In the case of a super-natural interaction schema, abstract schemata are recombined in distinct ways to express a meaning. For example, the ability to see through walls encompasses various abstractions that are related to our real-world experiences and reality-contradicting ideas.

In some instances, super-natural interaction can be combined with fictitious narrative elements. For example, in a flying carpet simulation [Pau+96], super-natural interaction techniques can be used to control a flying motion, whereas fictitious narratives, in this case, based on Arabian folklore and movie adaptations, provide an additional layer to the application. In this particular case, the concept of a flying carpet can be considered to enable novice users to instantly retrieve declarative knowledge that describes what the application is about by activating application-specific expectations and inhibiting fundamental concepts about reality that are contradicted in this environment, for example, gravity. This intentionality and meaning have to be coupled to the implemented interaction schema and associated physical actions and sensorimotor schemata. If the controls are easily and fully internalizable, the application can also be considered super-natural. The resulting interface corresponds to the definition for 'super-natural techniques' provided by Nabioyuni, in which "the designer typically puts a "story" around the technique and pro-

vides the user with abilities beyond real-world actions ..." [NB15, p. 5]. If the fictitious narrative elements are considered supporting schemata that successfully facilitate the acquisition of the interaction technique, these story elements can be considered to increase internalizability.

Figure 6.1: Internalizability (I) dimension.

Congruence Congruence is related to the interface not being "limited to physical reality" [Ste16b], or being "not limited to laws of physics or real-world constraints" [Lub18, p. 13], or (external) plausibility [Hof+20]. In most cases, an interaction technique with low plausibility would also lead to a low congruence. However, both concepts are not always the same. Plausibility assumes an objective and cognitive assessment of an interaction being coherent with physics and reality. In contrast, congruence refers to the subjectively perceived lived experience of enacting a specific agent-environment system and the resulting effects on the cognition and behavior of the user within the system. This intentionally approaches realism from a non-objectivist perspective to analyze factors that are not present in previously published work, such as the FIFA framework [MLP16].

Congruence indirectly measures realism by measuring the effect of using technology on the perception of ourselves and the involved patterns of self. Often, this involves conceptual knowledge about the possibilities of interaction. A virtual phone call in an IVE would not be considered magic despite exceeding the natural human capabilities to communicate. It is a usual way of communication in today's real world and, therefore, congruent with existing long-term patterns of self. As long as we are not able to just fly to a location in the real world on a regular basis, unassisted flying in an IVE seems like magic or not congruent to reality. It is also possible to think of interaction that is highly realistic but not congruent, for example, object interaction in an IVE simulating a space station. In such a scenario, interacting with floating objects or people perceived as flying would be highly realistic and fully coherent with physical laws. The congruence of this scenario, on the other hand, can be considered relatively low. This form of object interaction in such an agent-environment system is different from typically experienced real-world object interaction for most users, so both experiential congruence and agential congruence are low for most users. On the other hand, it is also possible to create an interaction that is not realistic but still, to some degree, congruent. An example is a hypothetical constraint flying carpet interaction technique that is limited to a flight height of 50 cm and manually steered along pre-defined paths. This locomotion technique would, on the one hand, have a low experiential congruence as this cannot be experienced in real life, and the fundamental physical law of gravity is violated. The agential congruence of this technique, on the other hand, can be rated medium to high, considering this technique corresponds largely to driving a car along streets in the real world, which is a long-term pattern of self that many users of an IVE presumably possess. For most constituents of the pattern of self, e.g., behavior and agency, only a negligible difference occurs, and only some aspects are largely different (for example, some aspects of this experience have not been observed before).

Abtahi et al. utilize the terms *reality-based* (no difference to real-world), *illusory* (not-noticeable modification), and *beyond-real* (substantially different to the real world) to describe modifications to movements in their framework[Abt+22]. In general, these categories can also be used to describe congruence (see Figure 6.2). A maximum congruence is reached when the VR-specific pattern of self is well-matched with other real-world pat-

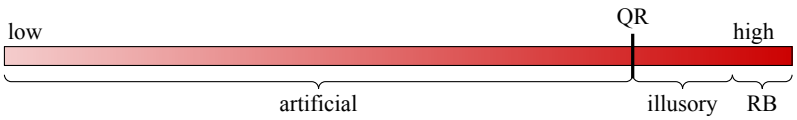


Figure 6.2: Congruence (C) dimension with labels adapted from the beyond-reality framework [Abt+22].

terns of self. Interaction with a high level of congruence can, therefore, be considered *reality-based* (RB)[29]. To a certain degree, the human brain cannot detect modifications to reality. In this case, interaction can be considered "illusory." If a certain amount of modification is exceeded (QR = "quasi-real"), the illusion becomes noticeable to a user, and the interaction can be considered beyond-real. Slightly enhancing the travel speed of a user in an IVE might not be noticeable [Ste+09], and making the employed pattern of self mostly congruent to the real world. After exceeding a certain detection threshold [Ste+09], for example, using the seven-leagues-boots-technique [IRA07], the deviation from reality becomes noticeable, and the interaction transitions from the illusory category to the beyond-real category. While Abtahi et al. focus on locomotion and movement in their framework, we propose that the same categories can exist for other deviations from congruence. For visual tasks, for example, cues can be designed to be not noticeable at all (e.g., slightly enhancing the colors of an object of interest to induce a shift of focus), or additional information can be presented as clearly visible input which would not be observed in the real world (e.g., visibly highlighting an object with a colored border).

In the design of interaction techniques, congruence shows some similarities to what Nilsson proposed as *perceived naturalness* in his PhD thesis [Nil15]. In his thesis on locomotion techniques in IVEs, *perceived naturalness* is described as "the extent to which the user's experience of walking through a virtual environment using a particular locomotion technique is mistakable for the experience of real walking" [Nil15, p. 37]. While he references the concept of interaction fidelity by McMahan et al. [McM11] and acknowledges the influence of *biomechanical*, *simulation* and *display* fidelity, he states this objective description "does not account for the degree of perceived naturalness produced by a given walking technique" [Nil15, p. 37]. Therefore, he also includes the "continuous experience" [Nil15, p. 37] of using a locomotion technique and perceiving a locomotion technique as "reminiscent of the perception of walking" [Nil15, p. 39] into the *perceived naturalness*. The concept of congruence extends this idea to include the phenomenological perspective created by the pattern of self that emerges from acting within an agent-environment system with multiple reference points of the lived experience of reality.

The term 'congruence' has been used before to describe the alignment between expectation and experience in the context of interaction in IVEs. Baños, for example, utilizes the term congruence in a general sense to express the alignment of virtual experiences and real experiences [Bañ+00]. Latoschik and Wienrich define congruence more specifically as "the objective match between processed and expected information on the sensory, perceptual, and cognitive layers" [LW22]. Although both these descriptions show some similarity, the term as used in this thesis has a unique meaning derived from the pattern theory of self and should not be used interchangeably.

[29] In Jacob's reality-based interaction framework [Jac+08a] technology is explicitly excluded which is not the case here.

Enhancement The central question for enhancement is how the use of the interaction technique changes the user-world relation. Drawing inspiration from the principles of organ projection and organ obsoletion, the power of an interaction technique can be related to the "average human capabilities" (E_h), and different levels of enhancement can be distinguished (see Figure 6.3): *diminish* (less enhancement than E_h), *support* (slightly enhanced abilities, which may not even be noticed), *strengthening* (distinctly increased existent abilities) and *enable* (allowing to perform completely new actions that exceed what is possible in the real world). Furthermore, a motivation for enhancing abilities can be to *restore* capabilities to the average human level [Whi+18].

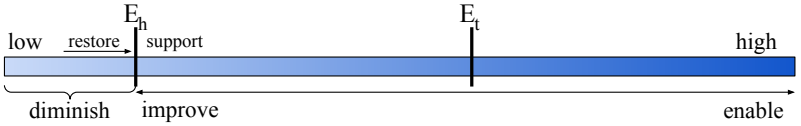
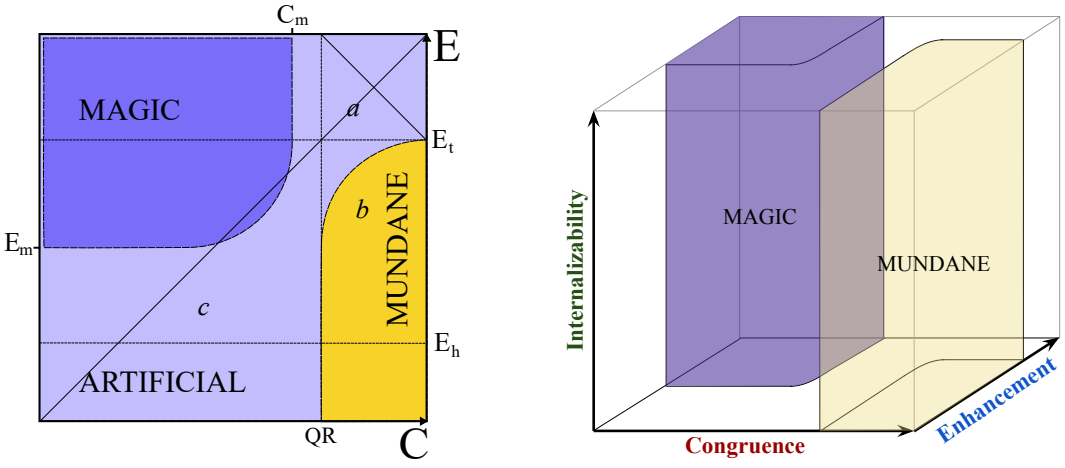


Figure 6.3: The dimension of perceived enhancement (E) with four levels of empowering: *diminish*, *support*, *improve*, and *enable*.

Arguably, even basic organ projections such as hammers (which enhance the ability of the first to hit) or scissors (which enhance the ability of fingernails to cut) [Kap18] can be considered enhancements to humans as a 'deficient beings' [GR88] that can only survive in the world by using tools and inventions. However, using a virtual hammer to hit a virtual nail would probably not be seen as particularly powerful by most users of an IVE and would, presumably, be considered a normal, mundane, or even natural interaction. In the same way, in which physical tools enhance our bodily abilities, providing additional sensory input, such as "heat vision" [Eri+19] or "x-ray vision" [Liv+13] can enhance abilities beyond our natural capabilities. Furthermore, cognitive processes such as memory retrieval [SH21a; Wil+21] and interpretation of the environment [Ito+16; Wil+21] can be enhanced by providing helpful additional information to a user.

IVE creates a technological medium in which interaction takes place. To some degree, enhancement can be quantified depending on the context of the interaction technique. For example, Nabiyouni et al. include the following factors in their testbed evaluation for locomotion techniques: accuracy, speed control, movement speed, spatial awareness, user comfort, user experience, fatigue, ease of learning, ease of use [Nab+15]. In the ICE cube framework, ease of learning and ease of use are, however, excluded from potential enhancement factors as they are related to the internalizability dimension. For quantifying enhancement, it is important to consider (i) which interaction technique is used as a reference, and (ii) if enhancement is achieved regarding an interaction in the real world or an interaction in the IVE. In this context, the first-order relation corresponds to a human user within the IVE using the most basic implementation of all possibilities the technology offers. Second-order relation, on the other hand, corresponds to humans outside of the IVE without any technology beyond the most basic naïve approach that can be transferred from reality, for example, using a pencil to write an annotation or using a hammer to drive a nail into a wall. Both involve necessary embodied technology that makes the task feasible at all. There are, however, some types of interaction with no counterpart in reality. For example, deleting or creating objects are unique interactions that are only encountered in computer systems, with no direct corresponding real-world action. In such a case, a good reference technique would be a conventional metaphor as the simplest form of interaction technique to perform such an action. For creating an object, for example, a naïve approach would be to open a menu, select an object from a list, and place it within the environment.



(a) Congruence-Enhancement plane. (b) ICE cube volume.

Figure 6.4: Construction of a conceptual model for spatially locating interaction techniques.

6.2.2 Classification

The axes enhancement and congruence construct an orthogonal coordinate system in which mundane (high C, low E) and magic interaction techniques (low C, high E) as extremes in the magic-mundane continuum [SU94] are diametrically opposed (see Fig. 6.4a). The definition of *magic* implies that besides magic and mundane interaction techniques, other categories exist that provide high enhancement while being congruent to the real world (high E, high C) and interaction techniques that are neither particularly powerful nor familiar (low E, low C). In the first case, it can be further distinguished between enhancements to E_h that exceed what is possible with current technology (E_t) while being principally plausible (a) and those that are also achievable with today's technology (b). The second case describes variations of performing a task, which relates neutrally to E_h while not being congruent to reality (c). The categories a and c, as well as 'magic,' are sub-categories of imagination-based [Kul09] or artificial interaction techniques. To be considered magic, interaction techniques need to pass a certain level of enhancement E_m and also exceed a certain degree of deviation from reality C_m. Expanding the C-E-plane with the internalizability axis yields a three-dimensional coordinate system we refer to as **ICE cube** (see Figure 6.4b). Mundane and magic interaction techniques are visible as volumes for each category, which contain interaction techniques that are easy or hard to internalize.

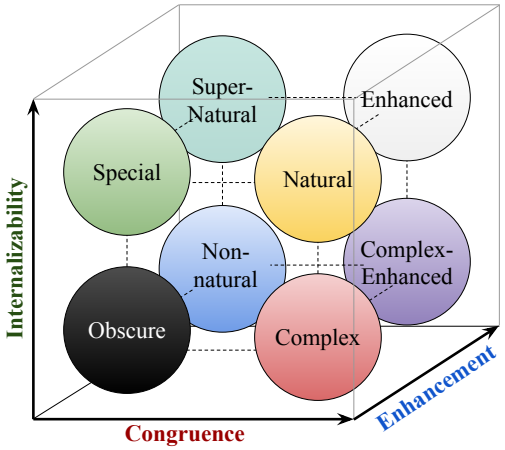


Figure 6.5: Location of different classes of interaction techniques within the ICE cube.

We assume that every interaction technique can be described using the properties 'congruence,' 'enhancement,' and 'internalizability.' Therefore, they can be located within the ICE cube, making the ICE cube a universal framework to describe interaction techniques. Locations in the volume (see Figure 6.5) represent a distinct combination of properties. We identify eight locations at the extreme points of the dimensions that can be utilized for classifying interaction techniques. To make the classes more graspable, we provide examples of locomotion techniques for each class in the following.

Natural (high *I*, high *C*, low *E*): interaction techniques that seem to a user like everyday interaction in reality can be classified as 'natural.' From the user's perspective, walking [Nil+18] or jumping without enhancements are 'natural,' as well as redirected walking [Raz05; Ste+09], considering that it seems like a congruent and non-enhancing way of locomotion to a user. Also, self-propelled simple vehicles [All+02] are mostly natural locomotion techniques (which are slightly leaning towards the enhanced class).

Enhanced (high *I*, high *C*, high *E*): Transferring real sporting devices to IVEs enables users to perform actions that are not possible in reality without technological support. Simplified virtual hang gliders [Soa+05; Glo+14], for example, enable average users, who are presumably not used to flying in reality, to fly with a reality-congruent device in an intuitive way.

Super-Natural (high *I*, low *C*, high *E*): Super-natural interaction techniques are highly internalizable magic interaction techniques. The term 'super-natural' has its origin in Nabioyuni's framework for locomotion techniques, which defines 'super-natural' as techniques that employ metaphors that show a low level of interaction fidelity [NB15], and in the definitions provided by Steinicke [Ste17a] and Lubos [Lub18]. Embodied flying [Liu+22], teleportation [Boz+16], and using a hoverboard [Sme+17] are examples of these metaphors, as well as interaction techniques without metaphors, such as walking on walls [Gie+18; Bec+19]. In contrast to using a hang glider, similar devices or technology, intuitive flying without supporting real-world flight devices [Tre+19; Pau+96; Med+19] can also be considered super-natural. Virtual jet-packs [SBH07], on the other hand, show more congruence than magic flying and are placed on the continuum between enhanced and super-natural interaction techniques, depending on implementation details and diegetic elements. 'Super-natural' is intentionally written as a hyphenated compound to emphasize its unique meaning and distinguish it from the commonly used adjective 'supernatural.' In terms of functionality, enhanced interaction techniques and super-natural techniques are closely related, as they differ primarily in their level of congruence. Since it can be assumed that the congruence of techniques changes as they become more established, it is also reasonable to expect that super-natural techniques, through frequent use and integration in everyday life, can eventually be considered enhanced. This is an important implication for hypothetical future aspects of the metaverse [DIG13], in which AR content and reality are blended, and interaction that would be considered super-natural in today's view can become a part of everyday life.

Complex (low *I*, high *C*, low *E*): In contrast to natural walking, some means of movement in reality require training to be successfully performed. Transferring these to IVEs creates complex interaction techniques. Examples are virtual wall climbing [Kos+17; Sch+19a] and, if physically correctly modeled and not simplified, virtual ice skating [Li22].

Complex-Enhanced (low *I*, high *C*, high *E*): In reality, flying can be performed using complex machinery such as airplanes, which require a considerable amount of training. Correctly simulating such a system in an IVE [YKT11; Obe+18; Vla+21; Aue+21] creates a congruent interaction technique that is not easily internalizable but provides the virtual working self-model of a user-enhanced agency.

Non-natural (low I, low C, high E): In contrast to super-natural interaction techniques, unnatural interaction techniques are magic, but difficult to learn. Performing teleportation by entering coordinates into command-line interfaces can be considered the artificial equivalent to super-natural point-&-teleport [Boz+16], as well as speech command implementations that follow a defined complex non-intuitive syntax.

Special (high I, low C, low E): Special interaction techniques serve a specific purpose without enhancing human abilities beyond E_h. Following the classification scheme, transferring the movement of fingers or hands to body movements [SCL17; Kim+08] or using semi-natural input modalities [Fre+20; Cho+22] without generating a feeling of enhancement can be considered 'special.'

Obscure (low I, low C, low E): interaction techniques of this kind are hard to learn and apply, and do not relate well to concepts of the real world. In their nature, obscure interaction techniques can mostly be considered a theoretical concept or an intentional design that aims to create an interesting, challenging, or fun experience instead of performing a task with a pragmatic reason. A 'loomesque'[30] obscure locomotion technique could be implemented as follows: Using different complex patterns of notes played on a flute, the user moves slowly in single steps in a specific direction. Other examples, outside of VR interaction, that can be considered obscure interactions are art installations that do not necessarily aim at providing an intuitively usable and effective interaction but, instead, challenge other aspects of human life and perception of self [Tri+20].

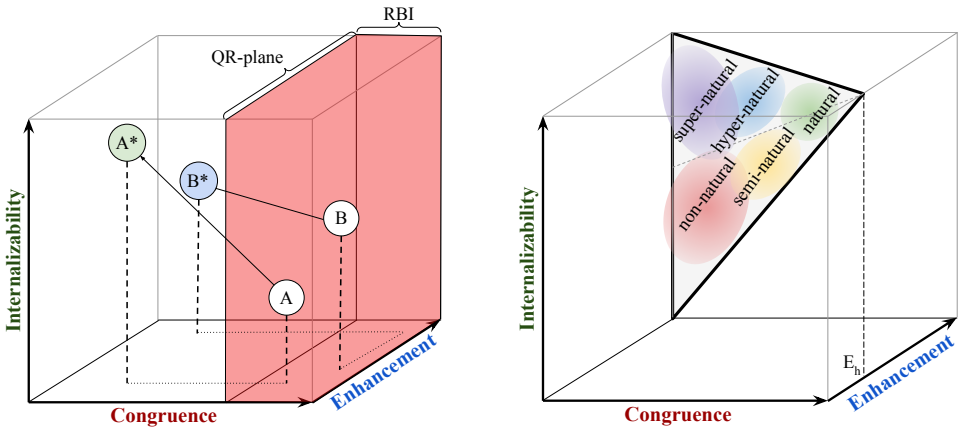
6.2.3 Discussion

We consider the ICE cube a product of the philosophy of instrumentalism in the sense that the model can be utilized as a possible approach to help resolve the identified use of heterogeneous terms regarding super-natural and magic interaction by providing a framework for a systematic description. The model is intended to provide a good balance between descriptive power and comprehensibility. The three-dimensional visualization can also help grasp the concept visually and facilitate discussions of spatial properties with peers. The ICE cube as a conceptual model allows for three main applications:

(1) As a framework to **support research** regarding interaction techniques by providing definitions and relations. Currently, it is difficult to retrieve all the literature about interaction techniques as keywords and search terms are either missing or not well-defined. The presented framework provides a reference for classification and terms and can be used to relate interaction techniques to similar approaches. The ICE cube complements other frameworks and offers an enactivism-based perspective that has not been adopted in previous research. Notably, the spatial arrangement, similar to the model proposed by Nabioyuni [NB15], enhances the interpretability of the concept and provides an intuitive approach to the developed ideas. This spatial visualization helps to clarify the relationships between different dimensions and facilitates a more accessible understanding of the conceptual ideas within the framework.

(2) As a framework for **interaction design**. Locating a specific interaction technique in the framework enables the structured application of a what-if-technique [Vid13] by independently varying the location on each axis. From such a process, trajectories emerge along which interaction techniques can be adjusted, for instance, by making a technique more powerful or more internalizable. Related techniques can be mapped onto these trajectories, making design decisions visually tangible. This approach enables designers to explore and compare different interaction methods systematically, allowing them to enhance or modify

[30] Loom is a point-&-click-style game published in 1990 in which a user performs actions by playing a corresponding melody on a digital flute.



(a) Jacob's reality-based interaction framework [Jac+08a] embedded within the ICE cube.

(b) Nabioyuni's framework for super- and hyper-natural locomotion [NB15] embedded within the ICE cube.

Figure 6.6: Combination of the three dimensions internalizability (I), congruence (C), and enhancement (E) into a cube model.

techniques based on their position within the framework. The visual representation of these adjustments can support the clarification of the tradeoffs and potential improvements, as well as potential flaws in interaction design.

(3) As a framework for **reflection** on one's own understanding of interaction techniques. The ICE cube proposes a perspective on the topic of interaction techniques. The selected enactive approach is not common in traditional HCI, however, it offers the potential to extend our understanding, regardless of whether the model is seen as entirely accurate or in need of improvement. As it is often said that all models are inherently flawed and only some are useful [Box76], our hope is that a multitude of perspectives can lead to a more comprehensive understanding by considering various aspects. The discourse on theoretical constructs and derived conceptual models is the core of philosophy and helps us in shaping our understanding and constructing our world.

One goal of the ICE cube is to incorporate and combine different views on interaction techniques and (non-)reality-based interaction. The ICE cube can be seen as a unifying construct that contains different models. For example, Jacob's tradeoffs in the RBI-framework [Jac+08a] can be visually interpreted as vectors (see Figure 6.6a) which enhance internalizability ($A \rightarrow A^*$) or power ($B \rightarrow B^*$) by moving interaction techniques out of the reality-based volume defined by congruence. The idea of increased ease of use and enhanced abilities corresponds to the concept of *tradeoffs* in Jacob's reality-based interaction framework [Jac+08a] (see section 2.2.3), but it describes specific tradeoffs that both improve performance and can be easily learned and used. Nabioyuni's and Bowman's model for hyper-natural and super-natural locomotion techniques [NB15] can be interpreted as an embedded triangle in the ICE cube, and their classifications can be interpreted as areas on this triangle (see Figure 6.6b) if the intention to provide techniques for faster or more accurate locomotion is considered an enhancement.

6.3 Pilot Questionnaire

6.3.1 Motivation

To verify the concepts developed in the course of this thesis, a preliminary explorative factor analysis (EFA) [Mat10] of ratings of questionnaire items that are intended to mea-

sure the three distinct dimensions 'internalizability,' 'congruence,' and 'enhancement' was conducted. The goal is to verify whether external experts in the field of HCI rate items coherently according to the proposed latent dimensions in the ICE cube model, and if the results enable a classification of techniques based on their location within a three-dimensional coordinate system using real data gathered from subjects who were not involved in the development of the presented theoretical concepts. Eight interaction techniques that are assumed to differ in terms of their internalizability, congruence, and enhancement are investigated. These techniques were deliberately selected to illustrate the orthogonality of these concepts and represent, to some extent, extreme cases of interaction techniques. While they may not have direct practical applications, they serve to highlight the distinctions between these conceptual dimensions. Due to limitations, this study can only be considered tentative, and further research is required to comprehensively analyze the proposed dimensions.

6.3.2 Study

Participants 15 participants (age M=29.6, SD=2.4) with an academic background in HCI and VR research (3 BSc, 11 MSc, and 1 PhD) were recruited for the study. All participants considered themselves experienced or experts in virtual reality interaction design.

Material and Methods The study was conducted in September 2023 as an online questionnaire. For eight different locomotion techniques, some well-known and some rather unfamiliar, the participants first received short textual descriptions of the interaction (see appendix D) and were instructed to estimate how the 'average VR user' would rate the presented statements after using the corresponding interaction technique. The following techniques were presented to the participants:

- *Teleportation*, similar to point-&-teleport [Boz+16] as a technique that is found in many applications and often considered super-natural.
- *Natural walking* as a natural way of locomotion that is directly adapted from reality. For this technique, it can be assumed that users already possess the required interaction schemata.
- *(Controller) Flying.* A locomotion technique that allows users to fly using the buttons on a controller.
- Users perform locomotion using a technique inspired by *finger walking* [Kim+08]. This interaction involves a simple schema that controls locomotion, however, it can be assumed that most users have not been exposed to this technique before.
- Users perform locomotion by typing commands into a *command-line interface.* While this technique is highly unlikely to be encountered in reality, it is an interesting case as it opposes the idea of System 1 cognition by being primarily dependent on symbolic interaction.
- Users fly using a realistic *airplane* cockpit simulation. Similar to a command-line interface, this interaction involves declarative knowledge.
- Users teleport using super-natural *orb portals* [HL19].
- Users travel with natural walking and an additional scaling factor for speed similar to the *7-league-boots* technique [IRA07].

Since there are no standardized questionnaires available to measure the three dimensions of the ICE cube model, an experimental questionnaire was developed for the three core concepts (see Table C.1). For each dimension, three questions were formulated that address

specific aspects of the conceptual background. The number of three questions per aspect was chosen to ensure that participants in the study would not have to answer an excessive number of questions, given the evaluation of eight interaction techniques. This limitation helps to prevent fatigue, which could otherwise distort the results. In the case of internalizability, the questions focused on the time taken to reach a state of mastery, i.e., the proficient use of interaction schemas, the complete internalization of these schemas, and the subconscious application in a System 1 manner. For congruence, the questions assessed both the overall plausibility and physical movements involved, as well as phenomenological effects, such as how the interaction technique feels to users. Finally, the questionnaire inquired about the effort required to use the interaction technique, the subjective experience of enhanced abilities, and the feeling of being more powerful when an interaction technique is used.

After reading the description, nine statements were rated by the participants on a 7-point Likert scale (see Table C.1). The score for each proposed latent factor was calculated from the rating of three manifest items per latent factor, rated from 0 (low) to 6 (high), and normalized to the range from 0 to 100, similar to the calculation of the overall score in the SUS questionnaire (see appendixF).

6.3.3 Results

Inter-Rater Reliability

To further investigate the similarity of ratings between participants, Krippendorff's alpha was selected as a measurement for inter-rater reliability. To recall[31], 15 raters with backgrounds in HCI and VR research took part in this study. The coding scheme was a 7-item ordinal scale presenting semantic differentials at each extreme (e.g., C1: "Completely different" versus "Identical") without intermediate descriptions. Raters did not receive further training or exercises concerning the rating. Krippendorff's Alpha was employed to assess the inter-rater reliability of the rating scheme [Kri18]. This statistical measure is particularly suited for studies with multiple raters and different levels of measurement. It is a robust measure that accommodates any number of raters, sample sizes, and missing data, making it appropriate for numerous study designs [Kri18]. In the present case, each rater independently provided ratings. The rated data were then input into the web-based statistical package K-Alpha Calculator [MBM24]. The analysis provided a reliability coefficient for the coding scheme, indicating the extent of agreement among raters beyond chance. The resulting Krippendorff's Alpha coefficient is 0.699 (95%-CI [.618, .752]). Recalling that the threshold for a satisfactory level of this coefficient is 0.80, as suggested by Krippendorff [Kri18], and an alpha value between .69 and .79 only allows for tentative conclusions [MBM24], the obtained alpha value can be considered to show a tendency towards a positive inter-rated reliability that needs further investigation.

[31] The remainder of this paragraph incorporates the template for reporting Krippendorff's alpha proposed by Marzi et al. [MBM24] with some adjustments for this particular study.

Factor Analysis

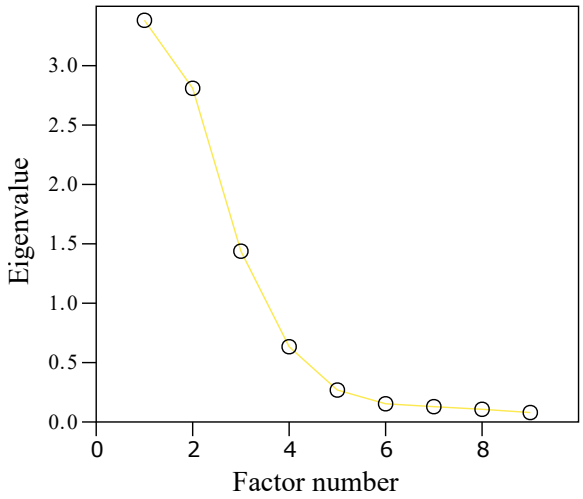


Figure 6.7: The scree plot for the EFA shows high eigenvalues for two factors and an acceptable eigenvalue (> 1) for a third factor.

To calculate the EFA, the software PSPP 1.6.2[32] was used. Overall, the dataset consists of 15 participants $\cdot$ 8 techniques $\cdot$ 9 items $= 1080$ data points (see appendix E). First, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was obtained. The calculated value of 0.75 indicated an acceptable variance for this preliminary stage of analysis, suggesting that the sample is adequate for factor analysis. Bartlett's Test of Sphericity ($\chi^2(df = 36) = 904.18, p < .001$) showed that the correlation matrix is sufficiently different from an identity matrix, confirming that there are significant correlations among the variables. With these criteria met, an EFA was conducted using principal axis factoring of the correlation matrix with varimax rotation. The number of three extracted factors was validated by examining the scree plot (see Fig. 6.7) and applying the eigenvalue-greater-than-one rule (also known as Kaiser–Guttman rule), which suggested three factors in compliance with the hypothesized three latent factors.

	Factor			
i	**1**	**2**	**3**	h^2
SSL	2.63	2.88	1.56	-
EV	29.2%	31.2%	17.3%	-
I0	**.94**	.05	.08	.89
I1	**.90**	.06	.15	.84
I2	**.93**	.08	.17	.90
C0	.05	**.87**	-.02	.76
C1	.11	**.97**	.02	.95
C2	.11	**.96**	.07	.94
E0	.03	.50	.24	.31
E1	.14	-.10	**.92**	.88
E2	.16	.00	**.77**	.62

Table 6.1: Rotated (varimax) PFA factor loadings, the sum of squared loadings of the rotation (SSL), explained variance (EV) of the investigated questionnaire items, and communality of the extraction (h^2).

[32] Available at https://www.gnu.org/software/pspp/.

Table 6.1 presents factor loadings for three factors along with the communalities (h^2) for various items, indicating a robust factor structure. For all items except E0, high factor loadings were calculated for the anticipated axis and low factor values for other axes (see Table 6.1 and Fig. 6.8). Items I0, I1, and I2 show high loadings on Factor 1, whereas items C0, C1, and C2 show strong associations with Factor 2, and items E1 and E2 are notably linked to Factor 3. The communalities h^2 range from 0.31 to 0.95, with most items displaying high communality, implying that the factors account for a significant portion of their variance. For example, C1 has an h^2 of 0.95, meaning 95% of its variance is explained by the factors. However, item E0, with an h^2 of 0.31, indicates a lower proportion of variance explained. Thus, E0 was excluded from the subsequent calculation of the 'enhancement.' The explained variance (EV) percentages show that Factor 1 explains 29.2% of the variance, Factor 2 accounts for 31.2%, and Factor 3 covers 17.3%, in sum, 77.7%. Overall, the high factor loadings and communalities suggest a well-fitting factor model that effectively explains a substantial portion of the variance in the data, particularly for items strongly associated with their respective factors.

Based on the factor loadings in the proposed model, the final scores for the proposed dimensions are now normalized as follows:

$$Internalizability = 100 \cdot (21 - I0 - I1 - I2)/18$$
$$Congruence = 100 \cdot (C0 + C1 + C2 - 3)/18$$
$$Enhancement = 100 \cdot (E1 + E2 - 2)/12$$

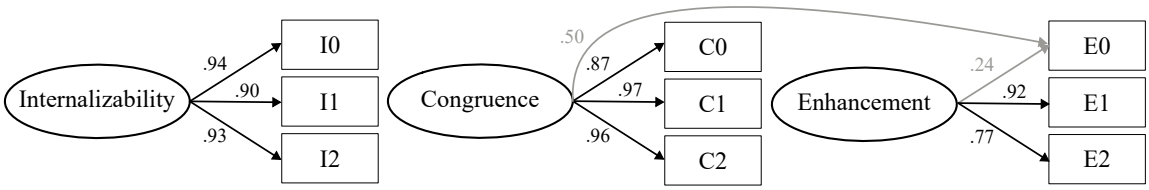


Figure 6.8: Ovals represent the latent factors internalizability, congruence, and enhancement. The rectangles represent the manifest questionnaire items. Only factor loadings relevant to the model are depicted (> .2).

For congruence and internalizability, scores range from 0 (low manifestation) to 100 (high manifestation). For enhancement, a score of 50 corresponds to no change in capabilities, values between 0 and 50 correspond to reduced capabilities, and values between 50 and 100 correspond to extended capabilities. The Pearson correlation coefficients for the scores were calculated as .159 (p = .083) for I and C, .274 (p =.002) for I and E, and -.015 (p = .867) for E and C, which indicates a low degree of correlation between factors.

Ratings of Techniques

Using the presented calculations, the scores for internalizability, congruence, and enhancement were calculated (see E). Boxplots were generated in RStudio 2023.06.2 to visualize the results.

Internalizability The scores for internalizability are spread across the full range. *Natural walking* and *7-league-boots* received the highest median scores (100 and 89). *Teleportation* and *Orb teleportation* received medium-high scores (78 and 61). *Controller flying* and *finger walking* received medium median scores (56 and 61), and *CLI* and *airplane* received very low scores (11 and 0). The interquartile distance per technique was similar for most of the ratings, ranging from 20 to 30 points. Only for natural walking and airplane, the dispersion was lower (3 and 11).

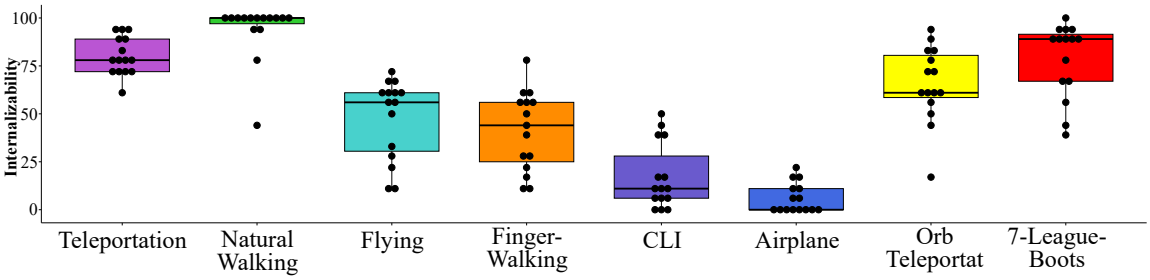


Figure 6.9: Visualization of the ratings for internalizability as box plots.

Congruence *Natural walking* and *airplane* received the highest median scores (100 and 89) regarding congruence. 7-league boots received a medium-high score (78). All other techniques received very low scores between 0 (*CLI* and *orb teleportation*) and 31 (*controller flying*). The dispersion for *CLI*, *teleportation*, *orb teleportation*, and *natural walking* was considerably lower (between 0 and 11) than for *controller flying*, *finger walking*, *airplane*, and *7-league-boots* (between 19 and 23).

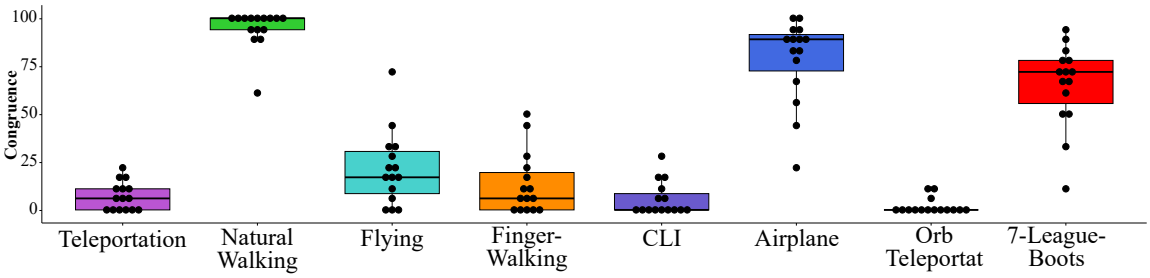


Figure 6.10: Visualization of the ratings for congruence as box plots.

Enhancement *Teleportation* and *controller flying* received the highest median scores for enhancement (both 75), followed by *airplane*, *orb teleportation*, and *7-league-boots* (all 67). *Natural walking* and *CLI* received medium ratings (42), and *finger walking* received a low rating (25). The dispersion of the scores per technique was considerably higher than for internalizability and congruence. In the case of *CLI* and *Aiplane*, the range of scores is very high (8 to 100 and 0 to 92).

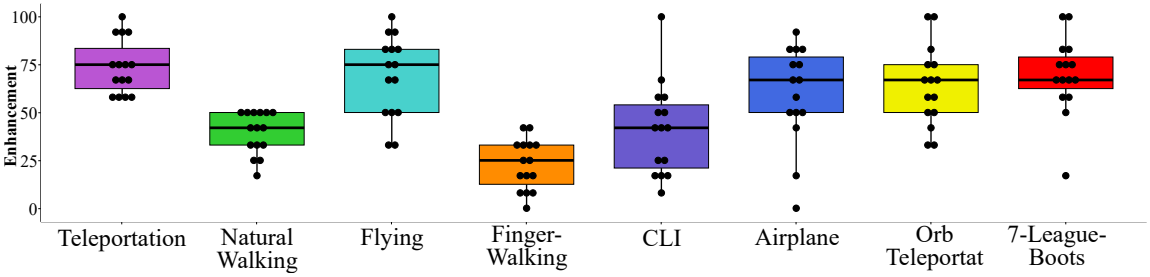


Figure 6.11: Visualization of the ratings for enhancement as box plots.

ICE cube The median scores and interquartile ranges can be used to visualize the location of every interaction technique within the three-dimensional ICE cube as volumes, with the median determining the location and the interquartile range determining the scaling per axis. The location within the ICE cube corresponds to a class of interaction techniques within the developed conceptual framework in this thesis (see section 6.2.2). Some techniques are located with a high degree of precision, characterized by a low interquartile

distance. Natural walking, for example, was considered a fully congruent, highly internalizable interaction technique, which, however, does not enhance human capabilities, making it a natural interaction technique. Point-&-teleport and orb teleportation can be classified as super-natural. The airplane cockpit simulation can be considered a complex-enhanced interaction technique. In the other cases, participants rated the interaction techniques less consistently, but their responses were still coherent enough to enable a meaningful classification. The location of finger walking and locomotion via a command-line interface corresponds to an obscure interaction technique, whereas flying using controllers would be considered a super-natural or non-natural interaction technique, depending on how easily the interaction schemata can actually be internalized in a practical application. The ratings for the 7-league-boots technique suggest a location between the classes natural, enhanced, and super-natural, depending on the actual threshold between the classes.

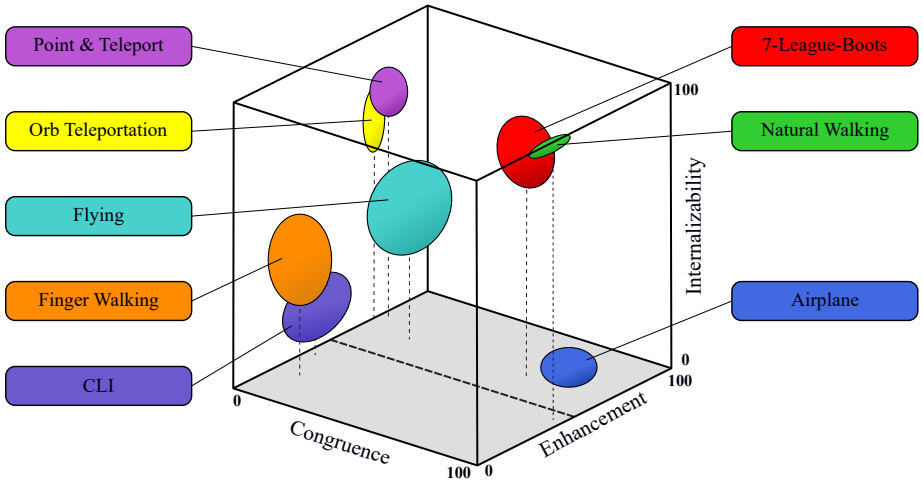


Figure 6.12: Evaluated localization of interaction techniques within the ICE cube based on the pilot questionnaire. The dimensions of each sphere correspond to the interquartile range of scores for each axis.

6.3.4 Discussion

The study results indicate that the concepts developed in this thesis,*congruence*, *internalizability*, and *enhancement*, can form a conceptual framework that enables a classification of techniques. The results show a substantial level of agreement among the survey participants, indicating that the approach can be considered successful. This consistency suggests that the model and the experimental questionnaire provide a reliable basis for evaluating interaction techniques.

However, considering the limitations of the pilot questionnaire, these findings should be regarded as preliminary, and they can only serve as a basis for formulating new research questions. The experimental questionnaire was chosen as a method that requires minimal resources to provide an initial evaluation of the framework, as experts were asked to imagine how the interactions would feel based on brief descriptions of each technique without any actual implementation or first-hand experience. This leaves significant room for interpretation that leads to increased variance in their estimations. For a more reliable exploratory factor analysis, a more sophisticated self-reported questionnaire with a large pool of initial items would be needed, and a set of diverse interaction techniques experienced by a large number of representative users in VR.

In this regard, MacCallum and Zhang suggest that the sampling size and recovery of population factors depend on the unique factor weights (communality) and the ratio of

variables to factors (overdetermination), and no "rule of thumb" exists that can generally be applied to every particular model and study [Mac+99]. As a reference, Baños et al. presented in the context of "presence and reality judgment" in IVEs (which in interaction techniques fundamental concept corresponds to the dimension of congruence) 77 questionnaire items to 124 participants in three scenarios (one per participant) to obtain factor loading for three aspects of virtual environments ("Reality judgment," "Internal/External correspondence," and "Attention/Absorption"), and still considered their study "small in size" [Bañ+00].

Some items show a greater interquartile range than others, which can likely be attributed to a larger room for interpretation in certain cases. This was particularly evident in the dimension of enhancement, in which a great variance of responses was observed. This may be due to the ambiguity regarding what the enhancement specifically referred to, as no concrete task was formulated. As a result, study participants may have had differing perceptions of how well a hypothetical interaction technique would perform in accomplishing a hypothetical task. This issue would largely benefit from a practical experiment.

6.4 Summary

In this section, a conceptual framework for classifying interaction techniques based on the three dimensions of internalizability, congruence, and enhancement was presented. A preliminary user study was conducted to assess whether experts who are unfamiliar with the conceptual foundations that led to the presented conceptual model would evaluate the techniques similarly. The results suggest that the framework yields reasonable results, with experts reaching comparable conclusions. However, a significant degree of variance remains, which requires further investigation in future research.

Since the pilot study produced fundamentally positive results, further analysis of the proposed dimensions could prove valuable. Due to resource limitations, the employed questionnaires can be considered a very basic form that can be improved in the future. Conducting a comprehensive and validated analysis of the scales was not possible in this work, but it represents a meaningful next step toward developing a reliable scale. This would help ensure more accurate and consistent measurement of the relevant dimensions in future research. It is likely that more accurate results could be achieved through a practical evaluation and first-hand experience of the interaction techniques rather than relying solely on the cognitive walkthrough method used in this study. Furthermore, the basic assumptions of the model can be examined in more detail in future research. For instance, the foundational assumption that internalizability and learnability, as well as realism and congruence, are distinct has not been further analyzed. The question of whether these dimensions are the most appropriate for such a model could not be conclusively resolved. However, the presented dimensions introduce fresh perspectives into HCI research, which can be considered a beneficial initial step for advancing the field.

The ICE cube aligns well with other proposed classification systems, such as those by Nabioyuni [NB15], Nilson [NSN16], and Jacob [Jac+08a]. Rather than introducing a fundamentally different approach, the model focuses on refining dimensions through the comprehensive analysis of enactivism-based concepts discussed in the previous chapters. By providing a spatial visualization, the ICE cube enables the comparison and categorization of interaction techniques, which offers an interesting approach for both analysis and ideation in the design and evaluation of interaction techniques.

PART III

APPLICATIONS IN IMMERSIVE TELEMENTORING

CHAPTER 7

PROJECT AND CONTEXT

In the first part of this thesis, conceptual aspects regarding super-natural interaction were analyzed and related to the enactive approach as a framework for cognition. Transferring the developed concepts and theoretical perspectives from a research context to applications in the real world can often be challenging. Still, an important aspect of super-natural interaction in IVEs is their application to solving real-world problems and their inherent implications, which are analyzed in this second part of the thesis.

This chapter presents heart transplantation as an important use case that can benefit from the potential of immersive technology and super-natural interaction. To systematically describe the use of novel technology in established environments as a transformative technological intervention that is aimed at improving existing ways of performing tasks, the ESTA framework is presented, which is derived from the concepts developed in the first part of this thesis. It extends the CHAT model [Eng01] to systematically describe and analyze human activities within their cultural and historical contexts.

7.1 On Transformative Technological Interventions

7.1.1 Motivation

The process of introducing new technology, such as devices, software, or processes, to existing fields of work can be described as a *transformative technological intervention* (TTI). In many cases, TTIs still present essential challenges [VB08], particularly when the technology disrupts established practices and workflows, even when a benefit for practices is, in retrospect, clearly visible [Nor98]. While engineers, designers, and researchers continually work towards a deeper understanding of HCI, not all developments can be successfully integrated into our technological and socio-cultural environment. The reasons for not accepting new technology are often multi-layered, interconnected, and, in some cases, not graspable at all.

The use of VR technology in real-world domains can still present challenges for engineers and designers [Gar+18]. In many cases, procedures and workflows have been established for many decades, and considering the young age of consumer-ready VR technology, researchers still have to find ways to integrate VR into real-world processes. Important factors that make VR systems usable in real-world domains are "simplicity, affordability, portability, and comfort" [Lan96]. Considering the form factor of current devices, such as the Meta Quest 2 and the HoloLens 2, great technical improvements have been achieved, but these systems are still rather cumbersome compared to fictional systems, such as the *Holodeck*. Like other computer systems, VR systems also often require specialized experts for maintenance, which inhibits their adoption in professional organizations.

Decision-makers, in particular, can be reluctant to accept VR technology, considering it gaming gear rather than professional tools [Lan96]. In 1985, Newell and Card describe substantial challenges regarding the application of science in technological environments emerging from the science being too low-level, too limited regarding scope, too late, or too difficult to apply outside of research environments [NC85], which appears to be often still

valid today. Often, the application of research is not only limited by what is theoretically possible, but also by what is accepted by users and what can be integrated into existing work and life structures. Perceiving the usefulness and the ease-of-use of a system are highly relevant factors for accepting new technology [VB08], but personal preference also plays a significant role [FAW19]. As Don Norman said, "[t]echnological change is simple; social, cultural, and organizational change is hard" [Nor98]. To explain, analyze, and mitigate various challenges in the introduction of new technology to existing workspaces, we present the ESTA (Enacted Self in Technological Activity) framework (see Fig. 7.1).

7.1.2 ESTA Framework

The ESTA model is intended to extend the second-generation Engström model for Activity Theory. It emphasizes the importance of tools in human-computer activities and their effect on the human user by including the contemporary philosophical perspectives of *enactivism* (in a liberal form which does not necessarily reject internal processing and representations [Ste14]), *pattern theory of self*, and *post-phenomenology*. These concepts share a similar focus on the context-dependent nature of human experience and challenge the traditional view of the self as a fixed and stable entity by emphasizing the role of context in shaping human behavior and experience. They also emphasize the dynamic and interactive nature of human experience and describe the self as an emergent phenomenon that arises from the interaction between the individual and their environment.

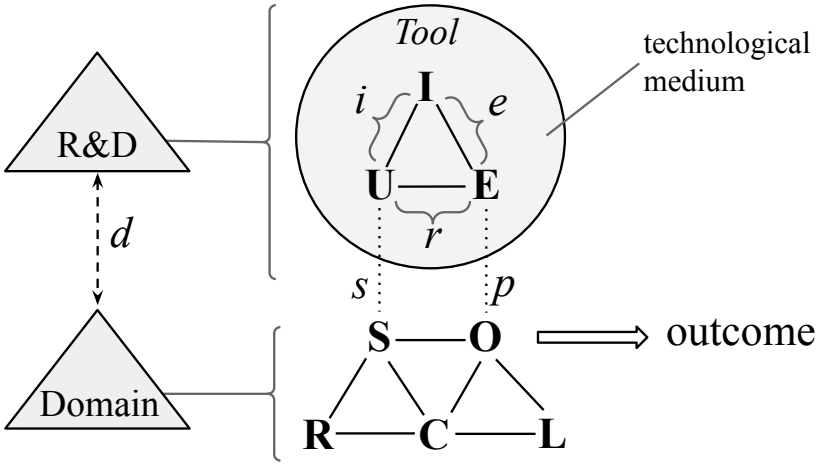


Figure 7.1: The ESTA framework is an extension of the CHAT model and Activity Theory in the context of transformative technological interventions. The subject S becomes a user U, a pattern of self, which is enacted during the technology-influenced activity. The object O is projected into a tool-dependent environment E. U and E are mediated by designed interactions I. The activity is embedded into a domain network of activities, whereas the TTI originates in an R&D process. Reprint from [Dew+23b].

TTIs are characterized by the introduction of novel technology to pre-existing fields of human life. TTIs are developed in R&D processes, themselves activities, typically carried out by interdisciplinary teams that include computer scientists, user experience designers, and engineers. The target activity, which the TTI is intended to improve, on the other hand, precedes the TTI and is embedded within a network of activities as part of a domain. This leads to at least two distinct expert groups: *technology experts* and *domain experts*. Both groups, with their diverse backgrounds, values, and intentions, form activities themselves that need to be synchronized to achieve a TTI that can be incorporated into the target activity. Strategies, such as user-centered, participatory, and user-generated design [San02] can help mitigate contradictions between these two groups (*d-contradictions*).

The *tool*, which is in many cases the focus of research in TTIs is of special interest in the ESTA framework. In general Activity Theory, the *tool* mediates *subject* and *object*, it defines how *subjects* can act upon their environment. With the introduction of new technology, the technological *tool* changes the *subject-object* relation and creates a new technology-dependent environment E. Through enaction within this environment, a new pattern of self – the user U – emerges which encompasses distinct properties in accordance with E: E and U are, in correspondence to the *subject-object* relation, mediated by an interaction I. This decomposition of T into E, U, and I allows for a more complex investigation of TTIs. The *tool*, as a result of an engineering and design process, defines U, E, and I, which form a *technological medium*. In this medium, the *dynamic structural congruence* [Mat02] leads to the enactment of a specific *tool*-dependent cognition. In this enactivistic view, it is important to emphasize that E, U, and I have to be seen as interconnected and interdependent aspects of a unified agent-environment system, and a strict separation is not always possible.

The specific technology-induced environment in which enaction takes place can vary. Traditionally, the environment of TTIs is a combination of artificially created entities embedded into the real world, for example, desktop computers in an office, smartwatches on our wrists, and mobile devices in our hands. These devices allow new interactions with the world and provide new affordances [PRL20; KN12] for interaction. The process of enaction and the emerging pattern of self depend on the structural coupling between the user's sensorimotor and cognitive capabilities and the features provided by the environment. In traditional HCI systems, the user exploits only a small subset of human capabilities and embodied interaction play only a minor role during interaction. During interaction with a graphical user interface, for example, the user consists only of "[one] hand with one finger, one eye, and two ears" [OI04], and the interaction is limited to only some modalities and specified ways of interaction. In some computer systems, an avatar is controlled by a user. In gaming, the choice of features of the embodied avatar has an influence on the arousal level and feeling of presence [LR09]. In VR research, the design of avatars is a major research topic, and changes in behavior depending on the embodied avatar, the so-called *Proteus effect*, have been described [KBS13; KSH19].

In contrast to interaction with the natural world (in the sense of the 'physical world without technology'), the enactivist sense-making in the context of technology interaction does not only encompass the dynamic interactions between the human user and the external environment, but also an intentional design and development process that creates the technological medium. This process is influenced by social factors, e.g., best or common practices. The technological medium with E, U, and I can, therefore, be considered a techno-cultural construct with no inherent way of interaction. Instead, medium-specific constituents regarding interaction have to become a part of the emerging user pattern. This can be described best with the term 'internalization,' both from the perspective of acquisition of cultural aspects of interaction in AT [KN97] as well as from the presented concept of internalizability presented in this thesis. In many cases, an intuitive way of facilitating internalizing the technological medium is to use standardization, which leverages previously internalized ways of interaction, for example, the WIMP concept (windows, icons, menus, pointers) in GUIs. Externalization, the opposite process of internalization in Activity Theory, is present in many systems in the form of tutorials and quick tips presented to novice users.

Primary contradictions in the sense of Activity Theory can occur in E, U, and I when the general user or goal requirements, values, technological possibilities, and context contradict each other. Often, tradeoffs have to be considered in the development process. Another major source of primary contradictions is errors in the development process and mistakes in

the design of the interaction, which lead to an inconsistent or error-prone system. Designers and developers have to match the environmental capacities and the user's abilities so that the intended effect of the TTI can be achieved. The technology-implied user pattern can also be considered from an ethical and inclusive perspective. HCI research is, as a western-dominated field of research, often *WEIRD* (western, educated, industrialized, rich, democratic) [Lin+21] and a narrowed view on the user may exclude certain groups or reinforce stereotypes [BL19; DEN16]. In many cases, the user pattern can enable an inclusive integration of various groups of users, or it can enable a neutral engagement without statements or prejudices.

By decomposing the tool node into the three nodes U, E, and I, new connections (s, p, t, e, i) and potential secondary contradictions can be described (see Fig. 7.1), which are of major interest for a prospective or a retrospective analysis. These new secondary contradictions between nodes in the ESTA framework can be characterized as:

Projection contradiction (p): A projection conflict occurs between the *object* of the initial activity and the environment created by the TTI. The environment includes some form of projection of the *object*, which is manipulated within the *tool*-dependent environment and which corresponds to the *object*. A p-contradiction describes difficulties in projecting between the activity's *object* and *outcome*, and the corresponding manifestation within E.

Relation contradictions (r): The relation between the environment and user is mediated by interaction with the intention of achieving a *tool*-dependent goal. When the user's capabilities do not match the environment's affordances, or when the environment does not seem to be part of a shared system, a relation contradiction occurs.

Internalization contradictions (i): The implemented interaction also needs to be internalizable by the intended group of users. Interaction can be multi-layered and vary within a single application, and designing interaction in such a way that a fast internalization of procedures is possible can be challenging. Especially when the intended group of users is not included in the development process, users may create a false mental model, leading to i-contradictions.

Efficiency contradictions (e): The implemented interactions have to be aimed at achieving *goals* while minimizing costs. Often, different ways of pursuing *goals* are possible, and developers need to find the most efficient way to reach *goals* within the TTI environment. These can differ from existing interactions between the *subject* and the *object* in the activity. Typically, the task-dependent efficiency of different tools can be evaluated in empirical studies, which are often based on laboratory studies to obtain generalizable results. In medical research, this measured performance under experimental conditions is referred to as efficacy, which is distinguished from the real-world observed effectivity and efficiency [BB20]. When the chosen form of interaction proves to be inefficient or when the findings from experimental studies cannot be directly applied to the specific context of the activity, an e-contradiction occurs.

Self contradictions (s): The enacted user pattern as a *short-term self-model* can conflict with the *long-term self-model* of a user. Both models rely on bodily, experiential, narrative, and cognitive constituents [Gal13]. The user pattern may adapt, omit, or even contradict constituents, which can lead to s-conflicts. In contrast to primary U conflicts, s-conflicts occur specifically between the intended user group and the general user pattern (which can work perfectly fine for other user groups).

7.1.3 Discussion

How does the presented ESTA framework differ from the established general model of Activity Theory? Do we really need another model in HCI? Depending on the approach and research focus, for example, from a sociological perspective, the abstraction provided by using only the concept of '*tool*' without further decomposition can be sufficient. However, the proposed decomposition may enable a more detailed analysis of implementation and integration issues of TTIs, which can be beneficial in the analysis and design of novel technology from the HCI perspective. Overall, we do not see the ESTA framework as a replacement for Activity Theory as a general model, but as an expansion suited to support HCI research with a strong focus on the phenomenological *subject-tool-object* relation. We believe it accounts for many challenges in describing technology-focused activities in the sense of contemporary waves [Bø06] and paradigms [HTS07] of HCI research. It also facilitates the expression of contemporary philosophical concepts in the context of HCI. Additionally, a combination with other models is not restricted by this enhanced view of *tools*. Furthermore, combining our framework with related models is easily possible, for example, third-generation CHAT (see Fig. 2.21c), or systemic-structural theory of activity. Central concepts can also be found in unrelated models, such as TAM3 [VB08], which focuses on the perceived usefulness and ease of use as factors for technology acceptance. Interestingly, despite being developed with a different theoretical background, these two factors correspond to the e- and i-contradictions described in our model (see Fig. 7.1). Cognitivist theories, which are still a dominant view in HCI, are in the same way not to be replaced by this approach, but are seen as an equally important complementary part of holistic research: While the ESTA framework can structure the analysis of TTIs by emphasizing the experiential factors of new technology, and providing insight in the relation of tools to activities, cognitivist approaches can be employed to allow an understanding of single contradictions by analyzing isolated factors. The presented ESTA framework itself can be viewed as a general tool for R&D activities, which can be beneficial for both analyzing and designing TTIs, and for communicating research findings or ideas.

The presented ESTA framework expands AT by drawing inspiration from contemporary views on human-environment relations: *post-phenomenology*, *pattern theory of self*, and *enactivsm*. These foundations share a common conception of the importance of context, mediation, and embodiment, and they can be clustered in the concept of *entanglement theory* [Fra19]. But are these theories useful in the context of HCI? To some extent, depending on the actual branch of enactivism [WSV17], the fundamental idea of enactivism limits or even rejects the main postulate of cognitivism that cognition is based on the internal processing of representations in the brain [Ste14], even though cognitive science has produced a large body of scientific work. Cognitivism is a dominant branch of research in HCI because of its success in describing and predicting human action in many cases. Furthermore, fundamental concepts of enactivism and its variants are still debated in the philosophical community (see, for example, [DJ16; VP21; MB22; War16]), and, at this point, it is not clear where debates will lead. We agree that a full rejection of traditional cognitivism is neither necessary nor possible at the current state [Whe14; MB22], especially in HCI, where symbolic interaction [Aks+09] is a fundamental concept for many types of HCI. Nevertheless, we argue that moving beyond the cognitivist framework broadens our understanding of HCI design and analysis. After all, discussing such fundamental topics is a vital part of progress in research. A liberal approach to enactivism, which emphasizes the non-representational nature in basic cognition but includes representation for higher cognition (which is not contradict the fundamental concept of autopoiesis [Zha21]), or related concepts, such as, embodied interaction, extended functionalism [Whe14] and distributed cognition are, from today's perspective, a good middle ground to shift the research perspective towards subjective experiential aspects without leaving established methods behind.

Similarly, both the *pattern theory of self* and *post-phenomenology* offer suitable approaches to shifting perspective from isolated, laboratory-based research to application-focused research, which can facilitate the deployment of TTIs in the real world.

We believe that the ESTA framework can provide a focused view for analyzing, describing, and understanding the challenges of human-technology interaction. The focus of the model is the decomposition of the *tool* used in an activity: A *tool* is not merely a technical development but a complex component embedded in human activities, which changes the relation of humans to the world. Introducing a *tool* leads to an enacted user self-pattern, changes in the environment, and individual modes of interaction, all of which must be considered in system design. In combination with Activity Theory, this interpretation of *tools* enables holistic research and development of TTIs. By including contemporary philosophy of mind and human-technology interaction, and linking theoretical thoughts to system design [HO17], it also offers an alternative practice-oriented systematic approach for conducting HCI research, which expands the catalog of previous models and methods.

7.2 Project Description

7.2.1 Immersive Telementoring

In this thesis, the use case of immersive telementoring in the field of heart transplantation is investigated. Since the late 1980s, telementoring has been used in various scenarios, often to enable video-based real-time collaboration between two hospitals or between one hospital and a remote location, such as battleships or mobile surgery units, to support real-time preparation and conduct of surgeries [RHG03; RYK07]. The general idea of telementoring is to support an operating surgeon, the mentee, with support from a remote expert surgeon, the mentor, using bidirectional audio and/or video real-time communication. In many cases, telementoring performs similarly to traditional on-site mentoring, and especially trainee surgeons benefit vastly in terms of operation times and surgery success, which implies that an application of telementoring is beneficial whenever on-site support is not possible [Err+19]. The application of MR technology in a surgical context has been a popular interdisciplinary research topic in medicine, computer science, human-computer interaction, and computer graphics for many years. In 1986, Roberts et al. Roberts et al. introduced one of the first medical AR imagery devices that superimposed information from computer tomography scans upon the surgery field as guidance. Over the last decade, VR and AR devices have advanced to a level of maturity that enables the practical application of this technology beyond single prototype devices. One such application is immersive telementoring (ITM), which exceeds the limits of video-based communication by providing features such as interactive virtual objects, realistic 3D scans, avatars, hand gestures, gaze tracking, video streaming and playback, and audio communication [Gal+20] to enable diverse ways of support of novice surgeons. In recent years, various immersive telementoring systems have been proposed that suggest benefits in this area or demonstrate the technical feasibility.

The STAR system by Rojas-Munoz et al. [RM+20a], for example, renders annotations and virtual tools in an AR head-mounted display, which are created by a remote expert who receives a video stream recorded by an overhead camera. In their research, they focus on analyzing the performance of surgical procedures using their telementoring setup in comparison to preliminary discussions of the planned surgery. Their research suggests that such telementoring systems improve task time and performance quality compared to audio-only communication [RM+20b]. The ARTEMIS system by Gasques et al. [Gas+21] utilizes multiple sensors to obtain a 3D reconstruction, which is annotated by a surgeon

in VR. These annotations are displayed in an AR HMD to the mentee. They focus on the quality of the 3D reconstruction, calibration procedures for hardware, and input methods for communication. In a similar way, Fischer et al. [Fis+22] analyze the streaming of point cloud reconstructions obtained by multiple RGBD sensors in surgical settings. They focus on the perceived realism of the 3D reconstruction. Research publications in the field of immersive telementoring share some similarities: Typically, they use an individually developed experimental prototype system consisting of specialized software and commercially available sensors, such as Microsoft HoloLens2, Azure Kinect, or RGB cameras. They differ in their research focus but typically include a technical part as an integral component of the TTI. The actual deployment in clinical procedures is yet limited, and the most common types of study are phantom experiments, simulator experiments, and system setups [Bir+22; PBC21]. Due to the critical situation in operations, the HoloLens2 moves only slowly from a proof-of-concept to an actual application in the operation room and clinical testing [Cas+21; Den+21].

However, deploying VR and AR technology outside of research facilities to support solving real-world tasks introduces challenges that need to be addressed. Often, applications provide a unique way of interaction, and the inherent required learning process disrupts the utilization of technology. Particularly surgeons, who are typically required to make fast and reliable decisions in stressful situations, simply do not always have the time to prepare the tracking hardware and remember unique interaction techniques for individual applications. To successfully introduce VR and AR technology to hospitals, it is necessary to provide means of input that reduce the additional workload on surgeons to facilitate an autonomous and competent use [Bur+19; Mac+14]. Highly internalizable super-natural interaction addresses these challenges and can be a good approach for interaction design for diverse tasks.

7.2.2 Use Case: Heart Transplantation

Heart transplantation (HTX) is both an important and challenging use case for immersive technology that may benefit greatly from the potential advantages of immersive technology due to the time-critical and logistical challenges of the procedures. The explantation of a donor organ can be performed in many hospitals in Europe (more precisely, the Eurotransplant region, which spans roughly from Croatia to the Netherlands). In contrast, only some specialized hospitals perform the implantation. The time span between the explantation of the donor heart and the implantation into the body of the recipient is only 4 hours to avoid increasing risks due to ischemia. The team of surgeons that explants the organ is typically sent from the implanting hospital to the remote location. They work in tandem with a second team of surgeons, who remain at the implanting hospital, and perform the heart implantation as soon as the organ arrives. Typically, heart explantation is carried out in parallel with the explantation of other organs, such as the liver, lung, or kidneys, limiting both the time and space for setup in the operating room. Additionally, the explanation takes place in a foreign environment in an external hospital. A crucial part of the transplantation is the evaluation and precise explantation of the donor organ at the explantation site. In most cases, the means for communication and synchronization between both teams are limited to occasional phone calls. A complete remote observation and support of the explantation is still not common today. In many cases, the explanting surgeons are less experienced and benefit from real-time communications with the implanting team.

HTX introduces unique problems that need to be addressed in a prototype system. In contrast to other stationary systems, no additional hardware and sensors can be utilized to prepare the operating room for recording the surgery. A connection to the Internet needs to be established without access points in the remote hospital. This can be achieved using

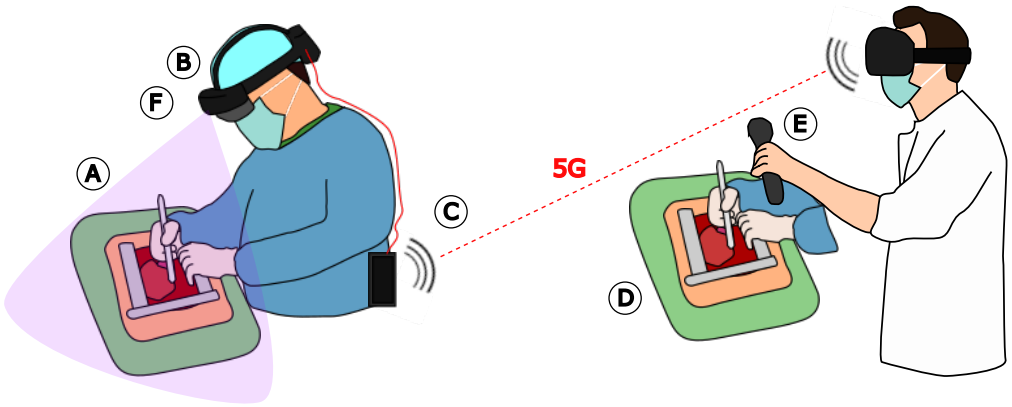


Figure 7.2: Planned pipeline of this research prototype. (A): Recording of situs using a single HoloLens 2 at the remote explantation, (B): Processing of the data on the HoloLens 2, (C): Streaming via 5G using a handheld smartphone, (D): 3D reconstruction of the situs at the hospital, (E): Annotations in Virtual Reality, (F): Display of Annotations at the remote location.

a smartphone as a 5G access point. Candidates for supporting the process of HTX with the help of mixed-reality technology are OST-HMDs such as the HoloLens 2. These head-worn devices can display diverse media in real-time and can be connected to the Internet to enable a remote surgeon to virtually join an explantation. One of the intended main applications of the HoloLens 2 is the support of workers by experts from afar using software such as the preinstalled Remote Assist in industrial cases. However, tests showed that this system is not suitable for surgeries, considering the accuracy of annotations and video recording angle, which is directed in a frontal direction instead of the downwards-tilted direction required in surgery. Furthermore, the transmitted video signal is only intended for a 2D annotation instead of an immersive application.

To extend the communication beyond video and audio, 3D annotations are a reasonable choice that has been implemented in previous projects. Research literature shows a sufficient accuracy for many surgical procedures with a deviation below 1 cm [Gal+20; Gsa+19]. With a correct 3D reconstruction, the transfer of 3D labels from the reconstruction to the real environment becomes possible. Different means of interaction can be implemented to enable annotations and other forms of communication, such as 2D drawing on a monitor or 3D drawing in virtual reality. In the case of open-heart surgery, the accurate placement of labels and annotations is, however, a more difficult task as tissue and organs constantly move or are occluded by the surgeon's hands or tools. From reviewing the literature and discussing existing systems with surgeons, we developed five key requirements that need to be met to allow for tests in real-world environments under clinical conditions and to be accepted by surgeons:

- **Safety**: The system must neither increase existing nor introduce new hazards to the patient during the operation. Furthermore, it must not increase the risk of surgical error due to distractions or malfunctions.
- **Sterility**: Interaction with the HoloLens 2 system during surgery needs to be fully sterile. The device must not be touched during surgery, and interaction with the HoloLens 2 is therefore limited to in-air hand gestures and speech commands.
- **Non-interference**: The setup should interfere as minimally as possible with the workflow of surgeries and tools of the surgeons, for example, individually fitted loupes. Surgeons should not have to adjust their postures to the device's requirements.

- **Ease of Use**: The startup and operation need to be as easy as possible. At the explantation site, only the head straps of HoloLens 2 need to be adjusted, which can easily be integrated into the *scrubbing*, the preparation process for surgeons before an operation. At the implanting hospital, surgeons should be able to carry out the setup of the VR environment on their own.
- **Mobility**: The setup is highly mobile and standalone, allowing fast deployment at various locations (within the hospital or outside). The operating room cannot be equipped with additional sensors. The HoloLens 2 and a mobile internet hotspot are the only devices for data acquisition, and no additional computer is used for further processing in the operating room.

7.2.3 ESTA Analysis

To systematically describe the challenges in the intended use case and the focus of this research, the previously presented ESTA framework is utilized. For a successful introduction of immersive technology in the context of HTX, all contradictions in the ESTA model need to be addressed. This is hardly possible in the technical development of an experimental prototype system. A practicable approach is focusing on specific key challenges and deriving specific goals for the development of a prototype system that demonstrate feasibility while not losing sight of the challenges that need to be addressed later.

Entities The first step is the identification of the entities that create the activity system. In our research project's activity, the *subject* of focus is a surgeon with the *object* of supporting decision-making during surgery from afar. The entities of this system are:

- **R&D** – *Research Activity*: A multi-disciplinary team of surgeons, developers, and HCI researchers that want to develop a TTI for immersive telementoring for heart transplantation.
- **Domain** – *Domain Activity*: The environment in which the TTI is intended to be deployed is hospitals.
- **R** – *Rules*: Safety standards and procedures dictated by law, the medical institution, and, to some degree, individual workflows in the hospital.
- **C** – *Community*: The hospital staff involved in the process of HTX.
- **L** – *Division of Labor*: The roles and tasks of involved people are clearly defined, e.g., the operating surgeon and the mentoring surgeon, and also supporting technical staff and manager.
- **Outcome** – *Activity outcome*: The outcome of the TTI is a new system that improves the process of heart transplantation
- **S** – *Subject* : In our research project, we focus on mentoring surgeons who provide support for remote surgeons explanting the heart.
- **O** – *Object*: The object is two-fold: i) the successful evaluation of the organ from afar and ii) the ability to communicate steps in the procedure to the remote surgeon in the form of annotations.
- **Tool** – *Technological medium*: An IVE. As proposed in the ESTA framework, the new *tool* is decomposed into an *environment* E, an *interaction* I, and the *user* pattern U. The specific properties of E, U, and I are based on design considerations and implemented in the R&D process.

- **E** – *Environment*: The situs as VR representation in the IVE as projection from the real-world explantation site to the IVE. The TTI replaces video-based streaming and phone calls with immersive technology to improve communication.
- **U**– *User*: The surgeon's avatar as a virtual representation of the surgeon, including extended aspects such as tools.
- **I** – *Interaction*: Super-natural interaction techniques that provide means to achieve the object of the activity (exploration and annotation).

Contradictions The second step is the analysis of contradictions. Possible contradictions we experienced during development are described and numbered in this section and depicted in the ESTA framework in Figure 7.3.

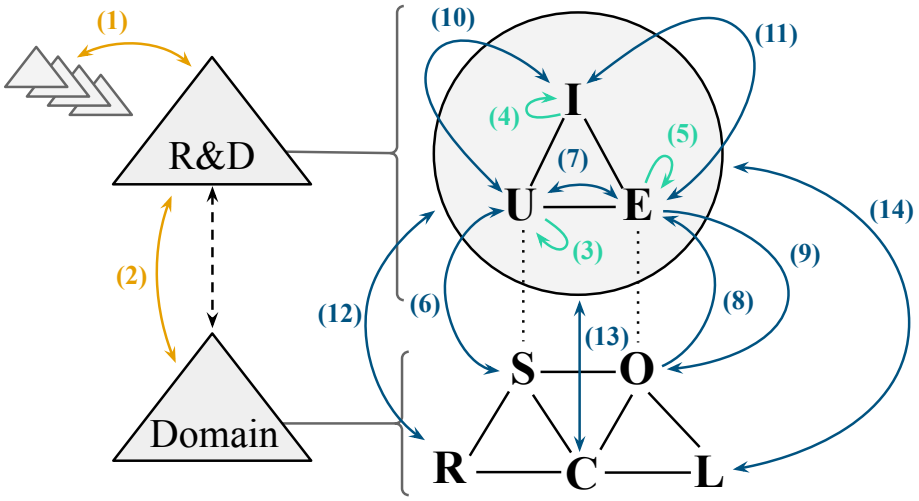


Figure 7.3: Identified primary (green), secondary (blue), and quarternary (orange) contradictions in our telementoring TTI depicted in the ESTA framework. Descriptions of contradictions are provided in chapter 7.2.3. Adapted from [Dew+23b].

(1) *Contradiction between neighboring activity systems*:
An explorative prototype system naturally conflicts with conservative activity systems that share the goal of ensuring the smooth functioning of the systems that are already in use within the hospital environment. The infrastructure in hospitals is focused on providing a reliable and secure service for existing systems. A big concern of TTIs in the hospital environment is safety and ethics. The TTI must not have a negative impact on the operation outcome, and the safety of patients has the highest priority. The decision process of allowing new technology can be very extensive, and risk management is necessary [Myt+10].

(2) *Contradiction between the domain of application and research*:
In our development, surgeons were members of the R&D team to include their expertise and insight into the process, and developers joined operation sessions to gain insight into processes and the environment where the TTI is intended to be integrated. The goal was to create a shared vision of the final prototype.

(3) *Contradiction within the TTI self-model*:
The surgeons need to enact a self-model for interaction within the IVE. It should be clear what the intention of the TTI is, so they are enabled to relate the different constituents to one another and understand the motivations for interaction. Contradictions occur when constituents of the patterns of self contradict each other.

(4) *Contradiction within the Interaction*:
The design of the interaction needs to be flawless, and there should be no errors in the implementation that negatively impact the interaction. A coherent system design and interaction style help users understand the functions and goals of interaction. Furthermore, the hardware used for interaction should fulfill certain requirements for accuracy and latency.

(5) *Contradiction within the projected environment*:
The environment should be capable of displaying the situs in real-time with sufficient quality and without artifacts that would negatively impact the evaluation of the organ. This is largely dependent on the selected hardware and implementation.

(6) *Contradiction between the self-model of the subject*:
The TTI creates a specific pattern of self that needs to be adopted by the surgeon. The constituents of TTI's pattern need to be designed in such a way that adopting the pattern is desirable and perceived as beneficial to the surgeons. In our development process, we observed that surgeons and medical experts are often reluctant to use VR technology, especially input devices, such as controllers, which makes the development challenging. Immersive technology requires, depending on the interaction design, some amount of training, which contradicts two important psychological needs of medical experts: the feeling of autonomy and competence [Bur+19]. Technical issues and work pressure also have a negative impact on job motivation [Mac+14], so creating interaction that reduces the possibility of user errors is necessary. Particularly in VR, the engagement with the technology can be disrupted [Kri05] when users blindly attempt to locate the VR controllers in the real environment or try to find specific buttons on controllers in the virtual scene. Using interaction only based on whole-hand input may improve this experience.

(7) *Contradiction between the TTI self-model and the projected environment*:
IVEs enable users to feel as if they were present at a different location [Sla09]. In our case, the surgeon should have the feeling of being on the explanation side to support their colleague. To achieve this, the capabilities and motivations of the user's self-model should match the created affordances and properties of the environment.

(8) *Contradiction from projecting the object to the environment*:
In our system, the *object* (supporting communication between mentee and mentor) is projected to the evaluation of the situs displayed in VR and the placement of annotations on the 3D reconstruction, the *environment* created by the TTI. This requires a life-like and real-time reconstruction of the surgical wound from data recorded in the operating room, which is suited for the evaluation of donor hearts and medical decision-making. The explanation site cannot be equipped with additional sensors, so all data has to be recorded using the HoloLens 2.

(9) *Contradiction in the back projection from the environment to the object*:
The communication is supported by annotations that can be displayed during the operation. The accuracy and visibility of annotations are crucial in avoiding errors. Additional communication channels, such as real-time voice communication and tracking of hands, can further enhance communication.

(10) *Contradiction between the TTI self-model and the interaction*:
The interaction in the TTI is an important part of the emerging pattern of self. Therefore, it needs to be designed in such a way that it forms a coherent system with other constituents of the self-model and supports the enacting of system-specific cognition and behavior. Interaction techniques should be targeted at the needs of the user group. In the case of this TTI, with surgeons as the primary target group, the interaction needs to be highly internalizable and self-explanatory.

(11) *Contradiction between Interaction and the projected environment*:
The interaction has to be efficient in solving the tasks. The two primary tasks identified in HTX telementoring are i) the evaluation of the donor organ and ii) the annotation of anatomical structures. The interaction needs to support both tasks with high efficiency. In the case of annotation, a central metric is annotation accuracy.

(12) *Contradiction between established rules and the TTI*:
Introducing new technology often requires high flexibility regarding infrastructure and can possibly interfere with the existing environment. Additionally, data security concerns prevent the introduction of TTIs that treat patient data in novel ways. Especially the USA Freedom Act (previously, the Patriot Act), which has been introduced by the US government to prevent terrorist acts by allowing investigators to access all online data stored on hardware in possession of US companies, is seen as not compliant with data security regulations – in particular the EU-introduced General Data Protection Regulation (*GDPR*, in German, *DSGVO*). This limits the use of services offered by US technology companies, which directly impacts possible technological approaches. The use of unknown devices, which may not be recognized as medical tools, such as the HoloLens 2, requires additional regulations. When a use in the operating room is intended, hygiene routines regarding disinfection and sterility must be developed to prevent infections. Clarifying the demands of regulatory entities and addressing the concerns is necessary to enable the use of immersive telementoring in real-life scenarios [Myt+10].

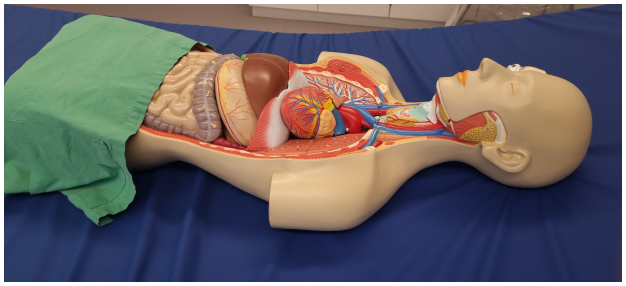
(13) *Contradiction between the community and the TTI*:
Both the social structure of surgeons and their education are highly hierarchically organized. Introducing new technology that challenges the traditional way can be particularly hard. To allow a practical use in the field of work and to secure funding for the development and introduction of the TTI, the ITM system needs to convince medical experts of its benefits regarding the treatment of patients [Key+18].

(14) *Contradiction between the division of labor and the TTI*:
The surgeons cannot be burdened with additional workload, so the system has to be designed in such a way that only minimal training is required for running and maintaining it, or new staff members can support the system [Key+18; Myt+10].

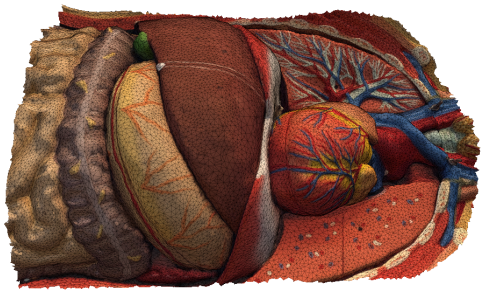
7.2.4 Goals of the Prototype System

The intended TTI of using immersive technology in the context of surgical telementoring yields a complex system with various challenges that have to be addressed to be deployable in real-life applications, especially from technical, design, and social perspectives. Addressing all of these challenges within a single project is hardly possible; therefore, it is crucial to focus on specific core elements while not losing sight of the overall system. Additionally, it is essential to ensure that no negative impacts are imposed on existing procedures, particularly in the healthcare sector. To largely mitigate risks during development (**11**, **13**), telementoring systems are often developed using phantom experiments or studies involving corpses [Bir+22]. We followed a similar approach to mitigate the challenges of hospitals as experimentation environments and minimize risks for patients by utilizing a representative model (see Fig. 7.4a) during development and testing. The model was created using the software Reality Capture [33]from 143 images captured with a Samsung S23 in a resolution of 6120 x 8160 pixels. The resulting 3D model is also used in chapter 9 and chapter 10 as reference material.

[33] Available at https://shop.scanner2go.de/products/reality-capture.



(a) A life-size anatomic doll (model: Erler Zimmer B235) used as a reference in the research and development process.



(b) Photogrammetry-based 3D reconstruction of the situs area. Medium-quality with 46k triangles.

Figure 7.4: The test setup for developing a proof-of-concept immersive telementoring prototype system and photogrammetry-based scan for research in VR.

To enable the use of this technology in hospitals, it is crucial to consider the method of deployment. In our experiments, we observed that doctors and medical professionals are not particularly enthusiastic about using VR technology. Additionally, there is typically no staff available to manage complex hardware systems. A promising alternative to these cumbersome setups are currently available mobile VR headsets, such as the Meta Quest 2, which require minimal setup time and support fully integrated, camera-based hand tracking. These devices offer a more practical and accessible solution for hospital environments.

Due to limitations regarding the resources in our research project, it was necessary to focus on some of the identified possible contradictions. The focus of development in our prototype system for HTX lies on capturing 3D data of an explantation site using only the internal sensors of the HoloLens 2 and streaming the data in real-time to the implantation surgeons using a 5G internet connection (**8**). On the organ implantation side, the data streams are merged into a 3D model that can be experienced in an IVE (**5**). Using Meta Quest 2 HMDs, the surgeons perform the task of evaluating the organ and annotating structure (**11**) by exploiting highly internalizable (**10**) interaction techniques based on whole-hand input specifically designed for this use case (**4**).

To summarize, the goals of the intended experimental prototype system for HTX telementoring can be defined as:

- **G1** Develop a mobile system that can be deployed in the operating room to project the situs to a virtual representation that is adequate for medical assessment of organ structure and motion.
- **G2** Capture the situs in millimeter resolution and life-like quality as a requirement for the subsequent transmission of annotations in medical telementoring.
- **G3** Create a system that facilitates a natural interaction between the clinical expert user and the virtual content in the IVE by implementing interaction techniques based on whole-hand input that are highly internalizable and highly efficient for the tasks i) situs exploration and ii) surface annotation.

G1 and **G2** are targeted in chapter 8, in which details of the prototype implementation are explained, and the system is evaluated by clinical experts. **G3** is investigated in chapter 9 (exploration) and chapter 10 (annotation).

CHAPTER 8

PROTOTYPE SYSTEM

This chapter presents the technical details of the HTX telementoring prototype system and approach, the challenges during development and their solutions. Based on a study conducted with heart surgeons who evaluated the results produced by the system the appropriateness of the system for supporting HTX is estimated. The system uses a single Microsoft HoloLens 2 to capture a point cloud of a surgery, utilizing only the built-in sensors, and streams the operation in real-time to a remote IVE.

8.1 Implementation

8.1.1 Hardware and Software

The HoloLens 2 is equipped with both an RGB camera and a time-of-flight depth sensor (AHAT). The AHAT sensor has a resolution of 512 x 512 pixels with a frame rate of 45 fps (reduced to 5 fps when hands are not present in the depth image), whereas the RGB camera offers various profiles based on application needs. The IMUs and optical sensors provide an accurate positional tracking that is suitable for medical and multi-user applications [Mat+21]. The applications on the HoloLens 2, HTC Vive Pro Eye, and Meta Quest 2 were developed using Unity 3D version 2019.4.34. For the HoloLens 2, the target build was set to Universal Windows Platform (version 10.0.18362.0) and ARM as the target processor to allow the integration of external ARM-based plugins. The HTC Vive Pro Application was executed on a workstation. In the case of the Meta Quest 2, the build process was performed by Unity to create and deploy .apk files on the device as a standalone application without a connection to a workstation.

8.1.2 Depth and RGB Recording

For a real-time recording of the situs using the HoloLens 2, two primary problems were identified: i) the device's recording direction deviates from the view direction of the surgeon. Only relying on the built-in sensors would require the surgeon to maintain an exhausting posture during surgery, which limits this approach. Furthermore, ii) the lighting conditions of operating rooms are well beyond the lighting conditions the camera are intended for, so they cannot automatically be handled by the sensors. To our knowledge, neither problem has been previously discussed in the research literature. We were able to find solutions for both challenges, which are presented in this chapter.

The HoloLens 2 is intended to seamlessly present virtual content superimposed onto the real environment in a small portion of the user's central field of view. This suits many scenarios, however, this is not the optimal configuration for surgical settings as investigated in this thesis. The direction of the built-in color camera is not suitable for recording surgical procedures, in which surgeons need to maintain a specific ergonomic posture with their head tilted slightly downward(around 20°to 40°, see Fig. 8.1). Surgeons may also wear individually fitted loupes that require a defined spatial relationship (angle and distance) between their head position and the surgical site, which has to be maintained while operating. When the color camera of the HoloLens 2 is aligned to the operation situs,

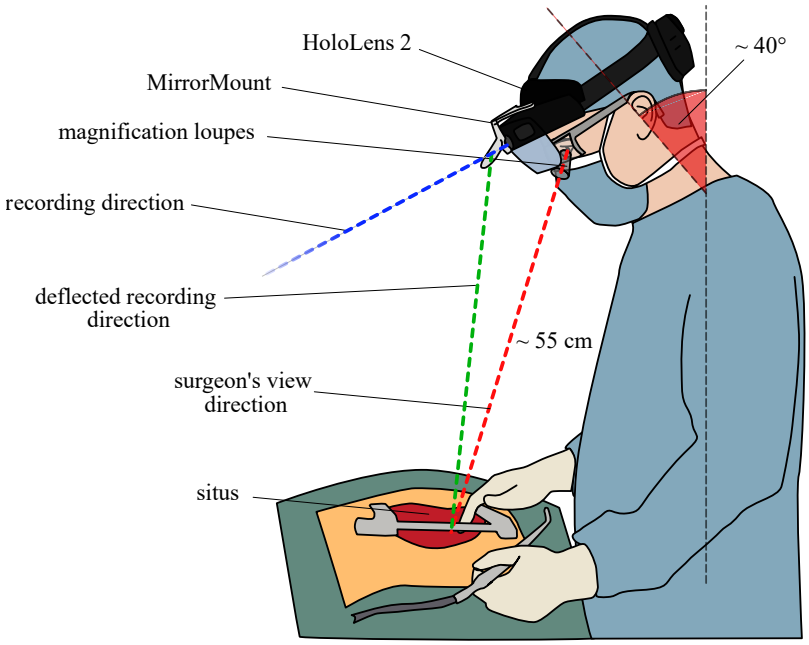
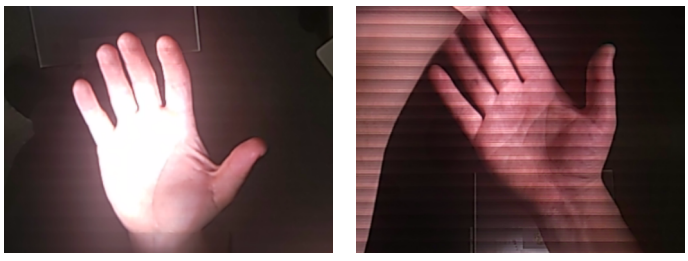


Figure 8.1: A surgeon wearing the HoloLens 2 with attached MirrorMount in the operating room in a typical posture during surgery. The camera's recording direction (blue) is deflected (green) to meet the surgeon's view direction (red) at the situs.

the surgeon is forced to tilt their head downwards at a much higher angle (around 50°to 60°), which is problematic from an ergonomic point of view. To address this limitation, a 3d-printed mountable mirror (in the following called 'MirrorMount'[34]) has been developed in the course of this project that modifies the camera's viewing angle, allowing the color camera to record the surgeon's hands and surgical wound during the operation. After several prototyping and test phases, placing a mirror with a diameter of 5 cm directly above the color camera and tilting it downwards by 35°was identified as a suitable spatial arrangement to focus the color camera on the hands of the surgeon. The mirror extends into the field of view of the depth sensor, causing reflections of infrared light emitted by the time-of-flight sensor. These reflections lead to incorrect depth information, affecting the 3D reconstruction of the HoloLens 2. To mitigate this, an additional cover is placed on the upper half of the depth sensor (see Fig. 8.6) to block the emitting IR light, reducing the depth image resolution to 512 x 320 pixels.



(a) Automatic exposure. (b) Visible rolling shutter.

Figure 8.2: Difficult lighting conditions in the operating room. Recorded with the non-modified HoloLens 2 RGB camera.

[34] The final version was mainly designed by my colleague Roman Bibo. Details regarding the design are published in [Dew+22a].

The lighting conditions in operating rooms are typically very bright to allow a good visual perception of the situs. Usually, focused operating room lights are pointed directly at the situs, which produces a steep slope of brightness. This can easily be handled by the human visual system, but it can be problematic for photoelectric sensors. Using the automatic exposure and ISO values from the H2 without modifications, the RGB images of the situs are overexposed (see Fig. 8.2a), and important features are lost. The exposure and iso values can be adjusted in the software using *Windows.Media.Devices.VideoDeviceController*-methods *ExposureControl.SetValueAsync(TimeSpan)* and *IsoSpeedControl.SetValueAsync(UInt32)*. However, due to the brightness of the LED operating room lights and the required short exposure time, the image shows rolling shutter artifacts (see Fig. 8.2b) when exposure and ISO are adjusted to correctly expose the situs. To obtain correct images of the situs, the HoloLens 2 has to be equipped with a neutral density (ND) filter with ND=3, which equals a transmission of 12.5% of incoming light.

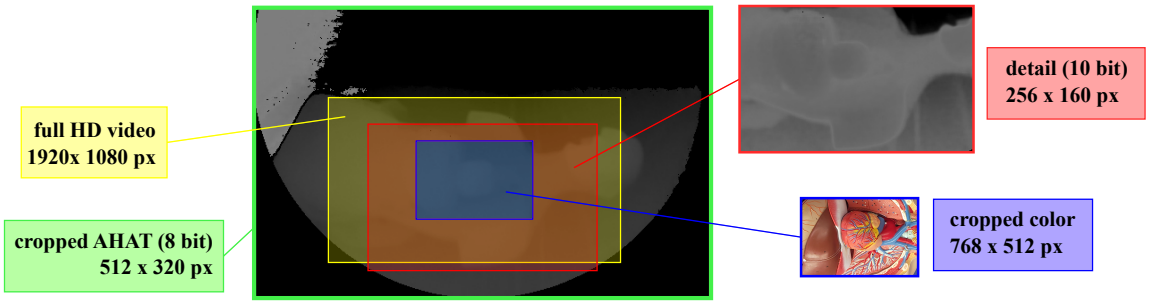


Figure 8.3: Captured data using the built-in sensors of the HoloLens 2. That AHAT sensor captures a wide field of view (green) that is encoded in 8-bit resolution. A quarter of this area, which corresponds to the position of the situs in the image, is encoded in 10-bit resolution (red). From the deflected full-HD video (yellow), a 768 x 512 pixel area is extracted for transmission (blue).

8.1.3 Video Transmission

The HoloLens 2 can establish a network connection to the remote hospital using the integrated WLAN adapter or via a USB-C cable. In our case, a Samsung S22 is used to enable the HoloLens 2 to access the internet via 5G and stream data to a remote location[35]. In experimental setups, the HoloLens 2 can also be directly connected to a laptop PC via USB-C. The package MixedReality-WebRTC 2.0.2 [Mic22] was integrated to allow for an easy implementation of video streaming. WebRTC has already been successfully utilized in comparable telementoring systems for real-time communication (For example, [RM+20b; Gas+21]). To connect the Unity application on the HoloLens 2 with the hospital-based workstation, a TURN server was implemented in the Unity application on the workstation PC, which enabled the negotiation and exchange of information between both clients via the TCP/IP protocol. This solution aims to establish a connection within a local network for experimentation purposes. For hospital environments with highly sensitive data and strict security regulations, other means of establishing a connection and encrypting data have to be implemented to comply with data security requirements.

Two different video signals are transmitted from the HoloLens 2 to the remote workstation: the first contains depth data at a resolution of 512 x 320 pixels, and the second contains RGB data at a resolution of 768 x 512 pixels. Both signals are encoded in the YUV_420_888 video format, which converts RGB data into full-resolution luminance

[35] With a cable connection, the Samsung 22 charges from the battery of the HoloLens 2, making it, unfortunately, a non-practical solution for many real-world applications.

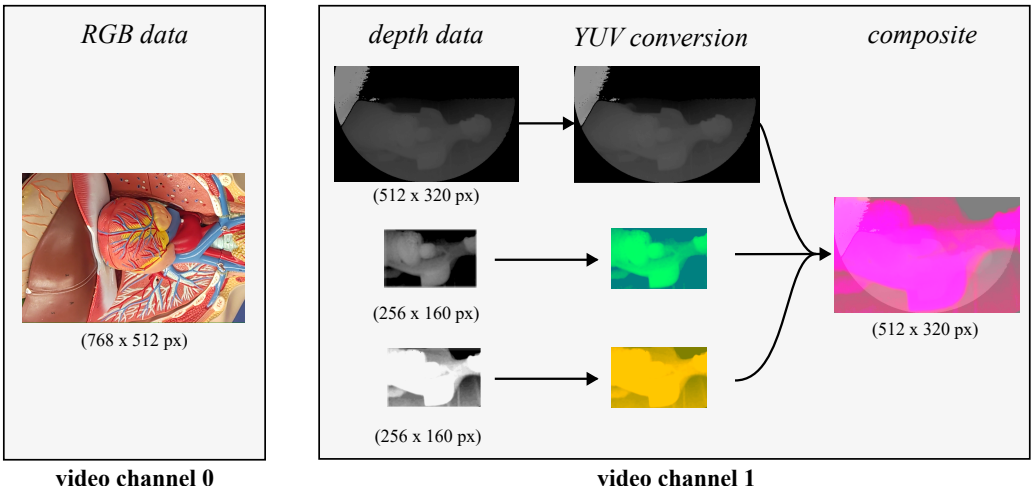


Figure 8.4: Video channel 0 (RGB) and channel 1 (depth) transmitted via WebRTC video streaming.

data (Y) and quarter-resolution chroma data (U, V) using the invertible transformations $Y = 0.299 \cdot R + 0.587 \cdot G + 0.114 \cdot B$, $U = 0.493 \cdot (B - Y)$, and $V = 0.877 \cdot (R - Y)$. The RGB signal is transmitted without further operations. The depth signal is split into three different signals with different resolutions (see Fig. 8.5). The first signal contains the full, low-resolution cropped AHAT frame (512 x 320 pixels) with values ranging from 15 cm (minimum range of the AHAT sensor) to 100 cm (maximum range of the AHAT sensor), represented by one byte as the Y-channel (depth resolution: 3.3 mm). The second signal represents the full-resolution detail view (with 256 x 160 pixels) in the area between a distance of 20 cm to 45 cm (1 byte for 250 mm, full resolution) as U-channel, and the third represents the same area between distance values from 45 cm to 70 cm (1 byte for 250 mm, full resolution) as V-channel (depth resolution: =1 mm). The three single signals are combined on the HoloLens 2 to form a single composite RGB video signal, which is transmitted to the hospital-based workstation (see Fig. 8.4). One frame is represented by a large array with 164 kB for the Y channel, followed by 41 kB for the U channel, and 41 kB for the V channel. Each frame is automatically encoded with settings determined by the WebRTC-based plugin, depending on the negotiation step between both clients and the available bandwidth. Typically, the incoming combined data rate of both streams after compression was measured at around 7 Mbit/s. After receiving and decoding frames from the depth video stream, the receiving workstation performs the described steps in reverse to obtain the original low-resolution cropped AHAT frame as well as both full-resolution detail frames. The final pixel value is determined by the signals in Y, U, and V. If possible, the full-resolution data stream is preferred for reconstruction; otherwise, the low-resolution signal is selected. The RGB video signal is received as it is. Both video signals are converted to RGB RenderTextures in Unity (the WebRTC plugin provides YUV video textures), which are then forwarded to the 3D reconstruction.

The transmission, as described here, can be considered sufficient for experimentation; however, the solution is not ideal and not sufficient for a real-world application. The encoding produces encoding artifacts in both video signals, such as plateaus for points with a similar distance to the sensor and additional structures (so-called 'mosquito noise') around sharp edges. In many actual use cases, such as real-time video conferences, this is an acceptable tradeoff between bandwidth, processing, and quality. However, in the medical domain, this is not acceptable when decision-making is based on possibly false values. Additionally, the synchronization of multiple video streams is not supported in the MixedReality-WebRTC package, so the depth and RGB streams have to be manually adjusted by setting an

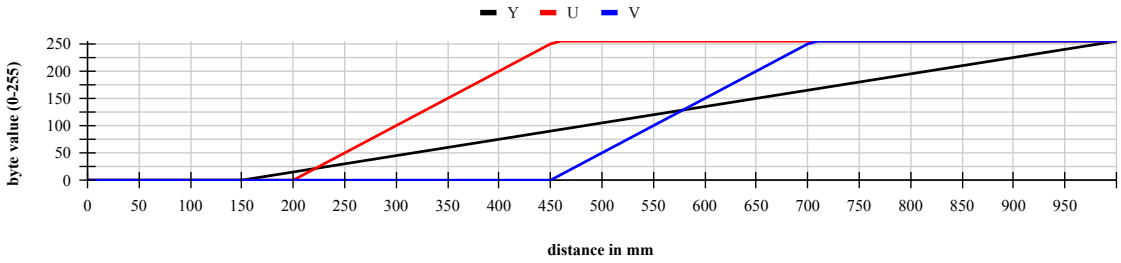
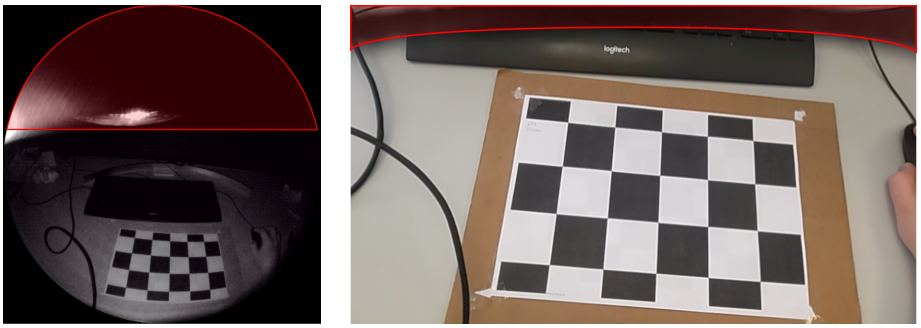


Figure 8.5: Mapping of distance values to byte values. The low-resolution image (Y) is transmitted with a depth resolution of 3.3 mm, whereas the high-resolution images (U, V) are transmitted with 1 mm resolution.

individual delay per stream for incoming frames. Typically, the achieved framerate lies between 20 and 25 fps, which can be seen as close to the lower threshold value for allowing remote surgeons to evaluate movement. Furthermore, MixedReality-WebRTC was marked as deprecated in 2022, so no further support for this package can be expected in the future, making the use of an alternative method for transmitting data necessary for future developments.

8.1.4 Camera Calibration



(a) Calibration frame from the AHAT sensor.

(b) Calibration frame from the RGB sensor (inverted along the y-axis).

Figure 8.6: Corresponding calibration frames of the AHAT depth camera and the mirror-deflected RGB camera. The red area marks which parts of the images are obscured by the MirrorMount.

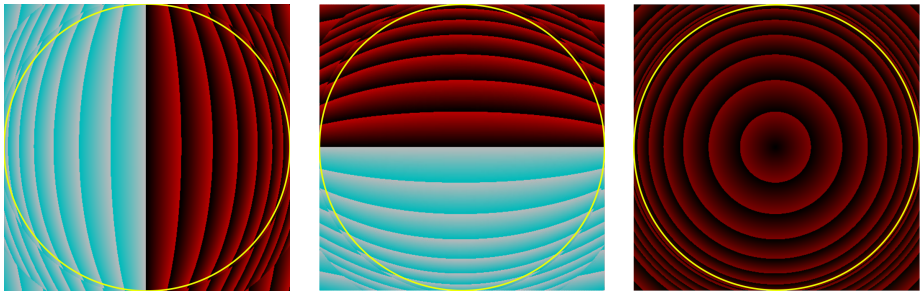
The 3D reconstruction requires a camera calibration in which the spatial arrangement of sensors and matrices for the mapping of color data to depth data is calculated. The calibration utilizes OpenCV's method *cv2.calibrateCamera* in Python, which assumes a pinhole camera model and calculates parameters from known 3D points in a planar calibration pattern, effectively eliminating the Z component [Zha00]. This approach simplifies the involved equations and enables a fast solution in which the homography is calculated, which describes a mapping between two planes. In a second step, the extrinsic ($[R\,t]$) and intrinsic (K) camera parameters are calculated from the homography. A complete calibration makes it possible to convert the depth data (X, Y, Z) into 2D image coordinates (u,v). For both the color camera and depth camera, the intrinsic and extrinsic matrices are obtained through the calibration of corresponding images with a 5 x 4 calibration pattern with an edge width of 30 mm (see Fig. 8.6).

$$K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}, \quad [R\ t] = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix}, \quad \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K\ [R\ t] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

With the calibration matrices obtained for both cameras, it is possible to convert the depth data of the AHAT sensor to color coordinates of the RGB camera using the following procedure:

- Let $[R\ t]_1$ be the extrinsic matrix for camera C_1, that transforms points in world space into the camera space of C_1.
- In analogy, let $[R\ t]_2$ be the extrinsic matrix for camera C_2.
- For $[R\ t]_1$, calculate the inverse matrix $[R\ t]_1^{-1}$.
- Calculate the relative transformation matrix $M_{rt} = [R\ t]_2 \times [R\ t]_1^{-1}$.

$[R\ t]_1^{-1}$ transforms points from the coordinate system of camera $C1$ to the world coordinate system. With $[R\ t]_2$, these points are transformed from the world coordinate space to the camera space of $C2$. M_{rt} combines this into a single matrix multiplication that transforms 3D points in the coordinate system of camera C1 to the coordinate system of camera C2. Using the intrinsic matrix K_2 the 2D points in the camera texture space of $C2$ can be obtained, which is used to determine the color of individual points.



(a) Horizontal. (b) Vertical. (c) Combined.

Figure 8.7: Visualization of the lens distortion parameters of the AHAT camera encoded in the LUT for the horizontal direction, vertical direction, and combined.

The HoloLens 2 does not support direct access to the sensor's intrinsic parameters (focal length, principal point) or the distortion coefficients. However, it offers functionality to project the depth 2D image points to 3D points with *MapCameraSpaceToImagePoint* and vice versa from 3D coordinates to the sensor image plane with *MapImagePointToCameraUnitPlane*. The latter function returns a pair of float32 values for each pixel in the depth image plane, which includes correction for distortion. To make these stored values usable in a real-time application, this method can be called for every pixel, and the result can be stored in an array with a size of the sensor's resolution (512 x 512) as a lookup table. The 2D lookup table can be encoded into a horizontal and a vertical texture, which are used in the reconstruction of the 3D information (see Fig. 8.7). In these textures, the barrel distortion becomes visible as contour lines. One contour line corresponds to a shift of 0.285 units, with the blue area representing negative values and the red representing positive values. Only the information inside the yellow circle has corresponding values in the AHAT depth image; the remaining information is not used in the reconstruction. The lookup textures are saved to a workstation or end device to reconstruct 3D data from the depth images during runtime.

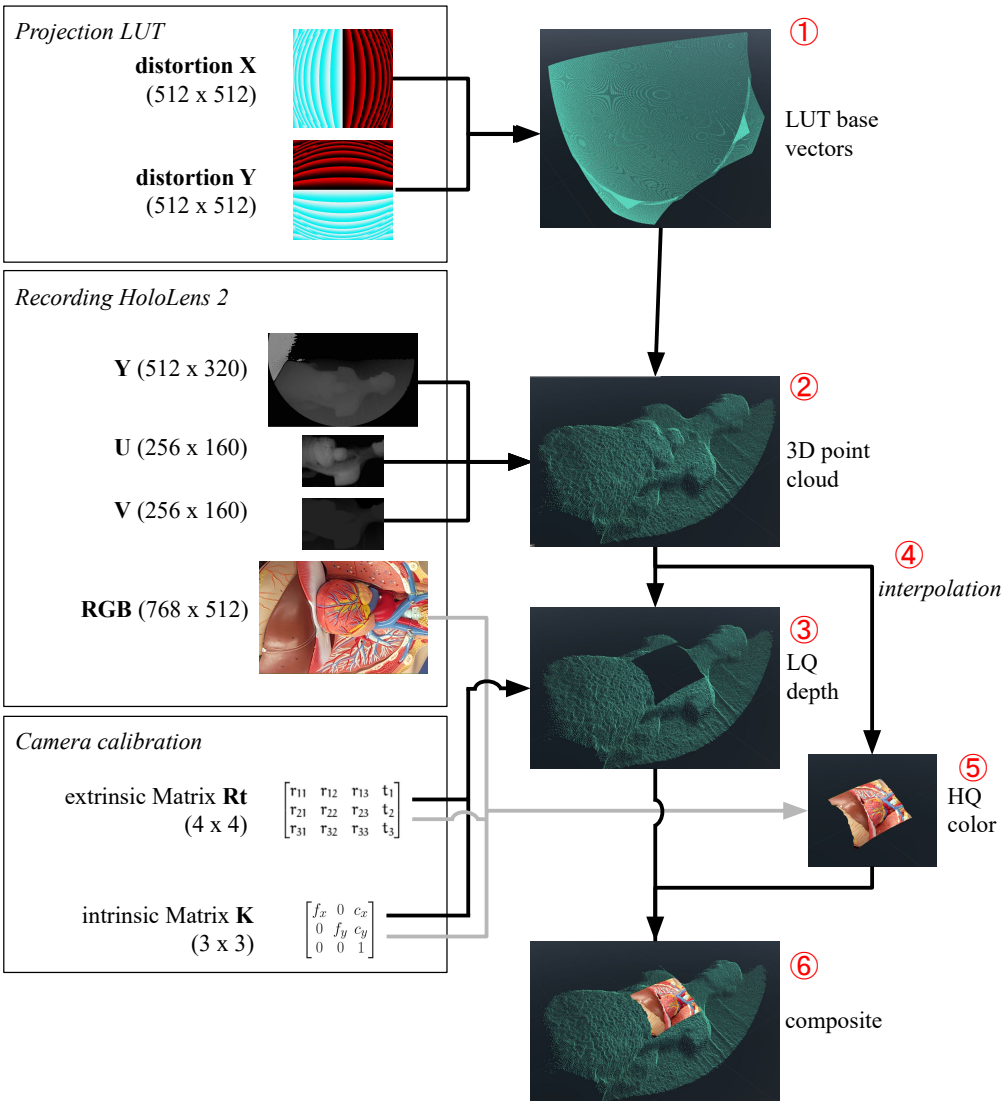
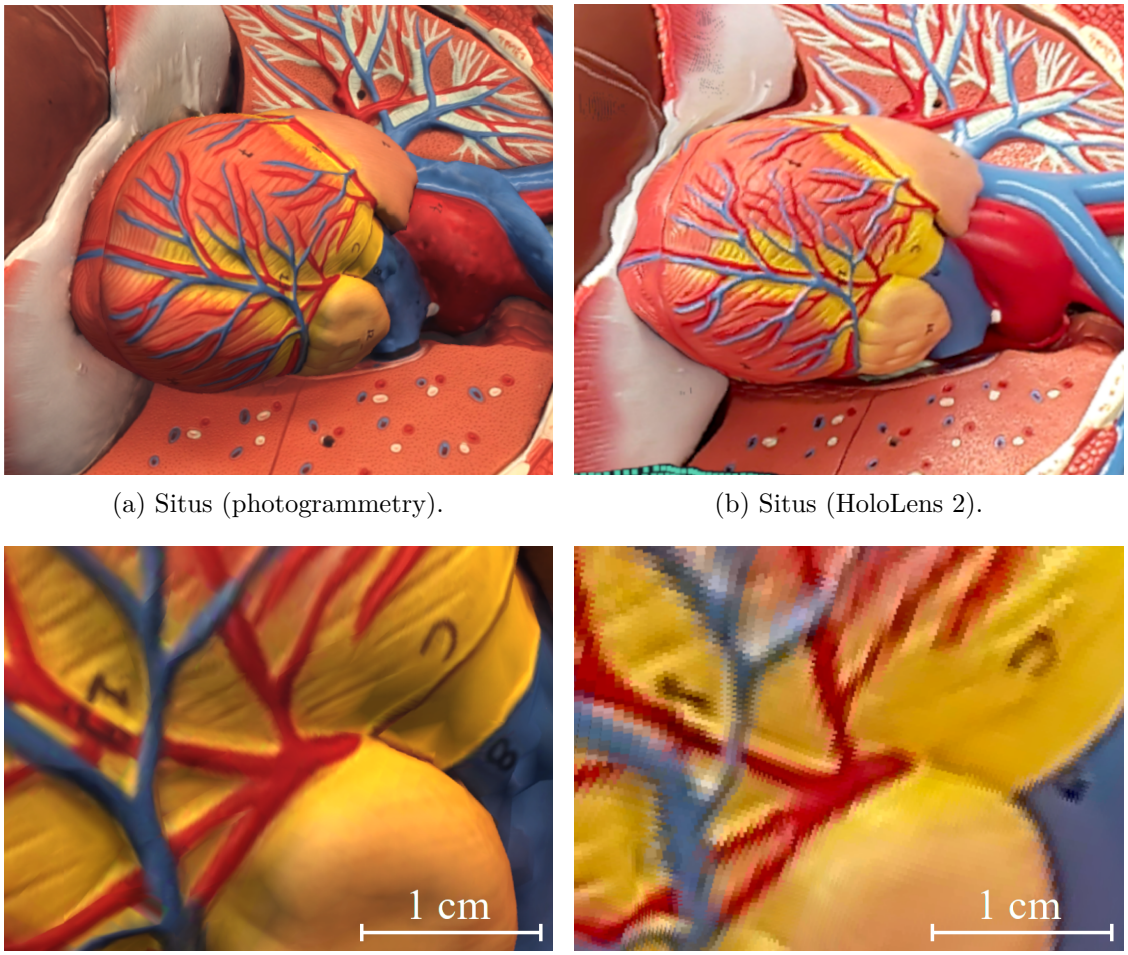


Figure 8.8: Steps in the point cloud reconstruction pipeline.

8.1.5 3D Reconstruction

To reconstruct the depth data, the distortion values d_x and d_y for a pixel with the coordinates (x, y) is expanded with a z-component of 1 to receive $\vec{d}_{LUT} = (d_x, d_y, 1.0)$. To obtain the 3D position $P_{X,Y,Z}$ from a pixel in the depth image $p_{x,y}$, the vector $\vec{d}_{LUT}$ is normalized and multiplied by the depth value. The reconstruction pipeline at the receiving end is implemented as a compute shader that receives the video stream textures, distortion parameter textures, camera calibration matrices, and other control parameters as input. The steps in the pipeline are visualized in Fig. 8.8. First, the normalized LUT base vectors are calculated from the distortion parameters (①). These vectors are multiplied with depth information transmitted in the YUV channels in the second video stream (②). From the low-resolution Y data stream, a low-quality reconstruction (8-bit) of the full AHAT field of view is created and colored in cyan (③). For the high-resolution (10-bit) data encoded in the U and V channels, the number of vertices is interpolated to increase the number of points at the situs' location (④). Using the static camera calibration parameters, the 3D points in the situs are used for a texture lookup to obtain a colored point cloud (⑤). Points that lie outside of the area captured by the RGB camera are colored in cyan. Both are combined to create the composite partially colored 3D point cloud (⑥).



(a) Situs (photogrammetry).

(b) Situs (HoloLens 2).

(c) Close-up (photogrammetry).

(d) Close-up (HoloLens 2).

Figure 8.9: Side-by-side comparison of a photogrammetry-based result and the 3D reconstruction generated with the proposed HoloLens 2-based system.

The overall achievable quality is limited by the utilized sensors. The spatial resolutions of both the RGB camera and depth sensor can be considered low for the intended surgical scenario, with only a small relevant section of the image (768 x 512 pixels in the full HD RGB image and 256 x 160 pixels in the AHAT image) for examining the surgical site from a distance. The 3D depth data shows some deviation from the reference photogrammetry-based model (see Fig. 8.10). This deviation primarily consists of high-frequency sensor noise and encoding artifacts, with an amplitude of less than 3 mm, as well as areas that are hard to record due to their orientation parallel to the camera direction. This makes the reconstruction of the situs sufficiently accurate to examine specific areas, however, fine details are lost, making the use of this system to assess fine details in heart surgery questionable. The color resolution can also be considered rather low. Compared to the photogrammetry model, fine details are lost, and the overall image appears pixelated in close-up (see Fig. 8.9c and Fig. 8.9d). With a higher distance, the difference becomes, however, less noticeable, and the impression is comparable to the photogrammetry model (see Fig. 8.9a and Fig. 8.9b).

A significant limitation of the presented approach is that the reconstruction of the situs produces good results only from a specific perspective. If the viewer's perspective deviates too much from the operator's original viewpoint, it becomes difficult to generate continuous surfaces, as the necessary details are not captured. However, since the operator's

perspective, from which the data is recorded, can be considered an important perspective for the mentor in the immersive telementoring process, this limitation is acceptable for a prototype at this stage. In the future, multiple AR devices or other sensors could be combined to jointly create a continuous reconstruction to improve the overall quality and enable viewing the situs from multiple angles.

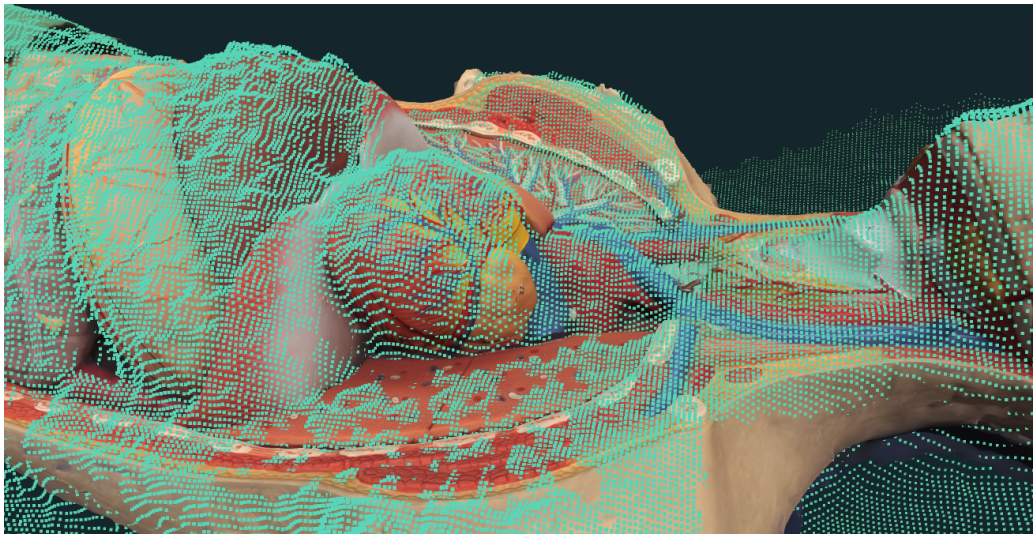


Figure 8.10: Reconstructed point cloud manually aligned with the photogrammetry-based 3D model.

8.2 System Appropriateness

8.2.1 Motivation

The presented system allows capturing authentic operations under real-world conditions. The mode of presentation is not limited to a 2D RGB video stream but also allows for a partial 3D reconstruction. The resolution is, however, limited by the built-in sensors of the HoloLens 2, so the question emerges whether the quality is sufficient for medical decision-making. To gain insight into this question, a small user study was performed with experienced surgeons who can compare the viewed material to their everyday work experience. Three variables were analyzed as important factors for telemedicine applications: the feeling of presence, with the two dimensions self-location (SL) and possible actions (PA), the subjective impression of media quality (MQ), and the medical appropriateness (MA). The mode of 'presentation' was analyzed as a factor on three levels: a 2D video display ('2D'), a 3D reconstruction viewed on a 2D screen ('3D'), and a 3D reconstruction viewed using a virtual reality ('VR') head-mounted display.

8.2.2 Study

Setup

For both the 2D view and the 3D model, an LG 49WL95C-WE monitor was used to display the content in a window with 1920 x 1200 pixels (WUXGA), which produced a view area with a width of 44 cm and a height of 28 cm. The participants were seated at a distance of 60 cm. For the VR presentation, a VIVE Pro headset was used to display a simple virtual scene (a simple floor and the colored point cloud as the only virtual objects). All

presentations were played on a Windows 10 workstation with an Intel i7 CPU, 64GB RAM, and a Nvidia RTX 3090.

Participants

We recruited 11 surgeons specialized in heart surgery (age: M = 36.9, SD = 5.5, 2 female and 9 male) from our hospital who have been performing surgeries for between one and 22 years (M = 11.4, SD = 6.0). One of the 11 participants was excluded from the study due to reporting difficulty 'seeing sharply' through the HMD, which may have been related to wearing glasses or a poorly adjusted or shifted device. Of the remaining 10 surgeons, five stated to have 'no' VR experience, three 'little,' and two 'some.'

Material and Methods

In all three cases, a comparable basic interaction was implemented to enable the examination of the reconstructed surgery. In the case of 2D, it was possible to move the video material on the screen by pressing the left mouse button, and it was also possible to zoom in and out using the mouse wheel. In the case of 3D, a similar interaction was used: The mouse wheel controlled the distance to the situs, and dragging the mouse while pressing the left mouse button rotated the camera around a point right in the middle of the situs. Although this interaction limits the possible viewing angles, it was preferred as no complex 3D navigation technique had to be learned by the participants, which could negatively impact the impression of the virtual reconstruction. In VR, the situs can be explored in a natural way by walking around the 3D reconstruction and moving the head closer.

The material shows a successful implantation of an LVAD (left ventricular assistance device) at the final stage of the surgery before the situs was closed. The HoloLens 2 was worn by a surgeon in his typical posture during surgery. The scene displayed the beating right ventricle of the patient's heart, a small part of the right lung, the diaphragm, transparent surgical drains, and the outflow track and driveline of the LVAD (see [Dew+22b] for details on the used material). For this stage of the prototype, we decided to show a looped 5-second clip of the beating heart without further instruments or actions as a baseline, as the most basic case of material that could be used in a telementoring process for evaluation of the situs from afar. The exposure and iso were manually set to ISO=100 and EXP=165750 to enable an examination of all areas of the situs. The material was recorded and played back using dedicated software in Unity. The received video streams (RGB and depth) were stored as lossless .tga files at a frame rate of 30 fps, along with the ROI as a byte array. For recording, the HoloLens 2 was directly connected to a Microsoft Surface Book 3 via a USB NCM connection using a 2-meter high-quality USB 3 cable to minimize transmission errors. The recordings were obtained using an early version of the prototype system, which transmitted the full RGB image at a resolution of 960 x 540 pixels.

To measure the feeling of presence, we used the spatial experience scale (SPES) [Har+15] questionnaire. It consists of two sub-scales, self-location (SL) and possible actions (PA), with four items each. For perceived media quality (MQ), no standardized questionnaire exists, so four items were added to account for different important aspects of media material in surgery:

- **MQ1**: 'The resolution and sharpness of the presentation were adequate.'
- **MQ2**: 'I experienced visually disturbing noise and artifacts.'
- **MQ3**: 'I was able to follow the motion in the presentation.'
- **MQ4**: 'The overall visual quality of the presentation was good.'

For medical appropriateness (MA), we integrated four statements that are relevant for telementoring in heart surgery:

- **MA1**: 'I was able to clearly distinguish between anatomical landmarks.'
- **MA2**: 'I was able to evaluate myocardial contraction and cardiac function.'
- **MA3**: 'The presentation allowed a clear impression of the situation in the operation room.'
- **MA4**: 'I would be able to support medical decision making from afar using this presentation.'

MQ and MA were developed with a senior surgeon at the hospital to incorporate medical expertise. All 16 questions were rated on a 7-item Likert scale labeled with 'fully disagree' (-3) at the left end, 'neutral' (0) in the middle, and 'fully agree' at the right (+3).

Procedure

The experiment was conducted as a within-subject design with the factor 'presentation' at three levels ('2D', '3D', 'VR'). The order of presentations was randomly assigned using a Latin square in 'Williams design'. Before the experiment began, the use case of telementoring was briefly presented, and important aspects of the presentation, such as media quality and medical appropriateness, that needed to be rated after each test run, were discussed. For each presentation, the participants were asked to freely explore the situs while having the telementoring application in mind. They were ordered to pay attention to medical details and image details that would enable decision-making from afar and to state what came to their mind ('thinking-aloud'). Whenever they felt confident enough to rate the presentation, they filled out a questionnaire. One experiment lasted typically around 10 minutes.

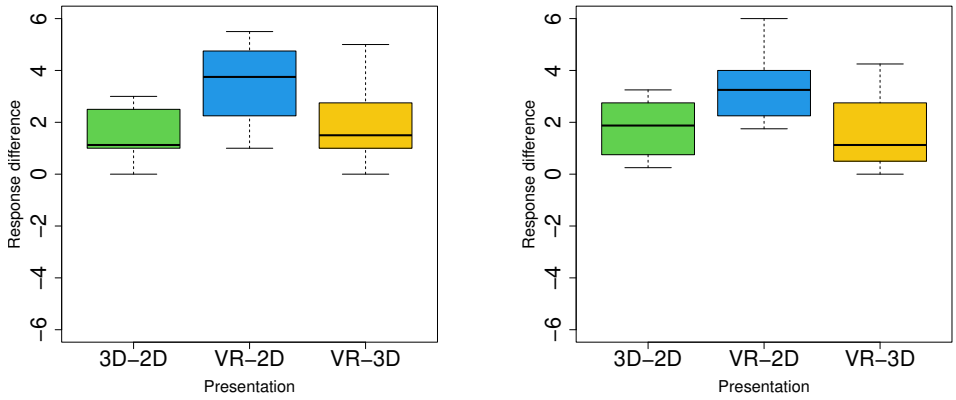
8.2.3 Results

	2D	3D	VR	3D-2D	VR-2D	VR-3D
SL	(-3, -1.75, +1)	(-3, +0.75, +2)	(+1.25, +2, +3)	(0, +1, +3)	(+1, +3.25, +5.5)	(0, +1, +5)
PA	(-3, -1.75, 0)	(-2.75, +0.25, +2)	(+0.25, +2, +3)	(+0.25, +1.75, +3.25)	(+1.75, +3.25, +6)	(0, +1.13, +4.25)
MQ1	(-3, +1, +2)	(-2, 0, +2)	(-2, +1, +2)	(-4, 0, +4)	(-2, 0, +5)	(-1, 0, +3)
MQ2	(-3, 0, +1)	(-3, 1, +3)	(-3, +1, +3)	(-1, +1, +6)	(-1, +1, +6)	(-1, 0, +1)
MQ3	(-2, +1, +2)	(+1, +2, +3)	(+2, +2, +3)	(-1, 0, +3)	(0, +1, +4)	(-1, +1, +1)
MQ4	(-3, +1, +2)	(-2, 0, +2)	(-1, +2, +2)	(-4, 0, +3)	(0, +1, +4)	(0, +1, +4)
MA1	(-3, 1, +3)	(0, +2, +3)	(0, +2, +3)	(-1, +1, +3)	(-1, +1, +4)	(0, 0, +1)
MA2	(-3, +1, +3)	(0, +2, +3)	(+2, +2, +3)	(-1, 0, +4)	(0, +1, +5)	(0, +1, +2)
MA3	(-3, 0, +2)	(-1, +1, +2)	(-2, +1, +3)	(-1, 0, +4)	(-1, +1, +5)	(-1, +0, +2)
MA4	(-3, +1, +3)	(-2, +1, +3)	(-1, +2, +3)	(-1, +1, +4)	(-1, +1, +5)	(-2, +1, +2)

Table 8.1: Descriptive statistics (min, median, max) for obtained data from the SPES questionnaire. 2D, 3D, and VR are the three levels of the factor presentation. 3D-2D, VR-2D, and VR-3D display the individual differences in rating.

Analysis of Measurements

The minimum, median, and maximum response values from the questionnaires are displayed in Table 8.1 (2D, 3D, VR) alongside the individual differences of ordinal categories for within-subject response between presentation (3D-2D, VR-2D, VR-3D). A Kolmogorov-Smirnov test revealed that, in almost all cases, answers on the SL, PA, MQ, and MA scales were not normally distributed. Therefore, the non-parametric Friedman test was selected to analyze differences in the rating of SL, PA, MQ, and MA between presentations with



a) *SL* b) *PA*

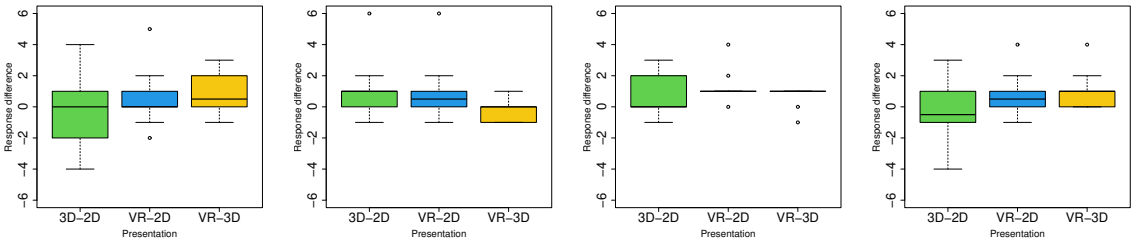
Figure 8.11: Distribution of differences of within-subjects ratings on the SL and PA scales.

repeated measurements. As a post-hoc test, Wilcoxon signed-rank tests were used for pairwise comparison. A Benjamini-Hochberg procedure [BH95] with a false discovery rate of $\alpha = 0.1$ was used to account for multiple comparisons.

SPES

The item responses on the SL and PA subscales were averaged to obtain a questionnaire value for PA and SL. Descriptive ratings for SL and PA are displayed in Table 8.1. Furthermore, the rating difference between presentations is displayed in Fig. 8.11. On the SL subscale, the Friedman test showed a significant difference between groups: $\chi^2(2) = 18.05, p < .0002$. Post-hoc test showed a significant difference for 3D-2D ($p < .017$), VR-3D ($p < .017$) and VR-2D ($W = 0, p < .012$). On the PA subscale, Friedman test showed a significant difference between groups: $\chi^2(2) = 19.1, p < .0001$. Post-hoc test showed a significant difference for 3D-2D ($p < .011$), VR-3D ($W = 0, p < .017$) and VR-2D ($W = 0, p < .011$). Comparing the median ratings on both scales measured in SPES, it follows that, in this setting, regarding the feeling of presence, a VR presentation provides the highest level of presence, followed by a 3D presentation, and last, a 2D presentation.

Media Quality



a) *MQ1* b) *MQ2* c) *MQ3* d) *MQ4*

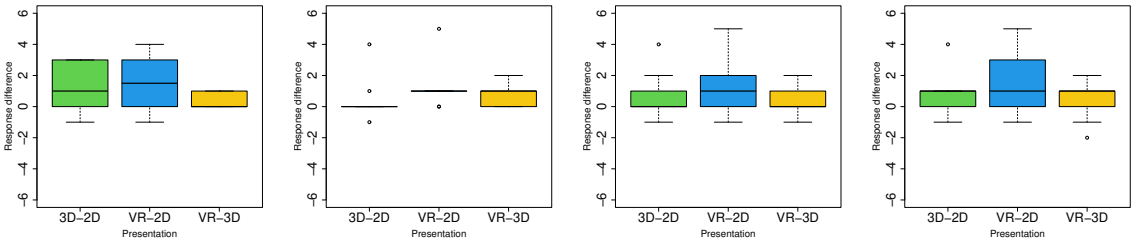
Figure 8.12: Distributions of difference in within-subjects ratings of MQ questionnaire items.

Items of the MQ scale were individually compared as absolute category response (ACR). The Friedman test showed no significant difference for MQ1 ($\chi^2(2) = 0.95, p = .62$), MQ2 ($\chi^2(2) = 4.2, p = .12$) and MQ4 ($\chi^2(2) = 3.95, p = .14$). For MQ3, Friedman showed a

significant difference ($\chi^2(2) = 10.05, p < .007$). Pairwise comparison showed a significant difference for VR-3D ($W = 5, p < .05$) and VR-2D ($W = 0, p < .05$). By comparing the medians of MQ3, the data suggest that the evaluation of motion benefits from a virtual environment, whereas 3D view and 2D video do not differ significantly.

Medical Appropriateness

ACRs of the MA scale were individually compared. Differences between groups are shown in Figure 8.13. Friedman tests showed no significant difference for MA1 ($\chi^2(2) = 5.85, p = .054$), MA3 ($\chi^2(2) = 2.85, p = .24$) and MA4 ($\chi^2(2) = 5.15, p = .076$). For MA2, the statistics showed a significant difference ($\chi^2(2) = 8.55, p = .01$). Post-hoc test showed a significant difference for VR-2D ($W = 0, p < .05$) and VR-3D ($W = 0, p < .05$). As the median of VR is higher than 2D and 3D, it follows that surgeons felt significantly more able to evaluate heart function in VR.



a) *MA1* b) *MA2* c) *MA3* d) *MA4*

Figure 8.13: Distributions of difference in within-subjects ratings of MA questionnaire items.

Additional Statements and Observations

Six of the 10 participating surgeons showed a positive reaction (e.g., exclaiming 'cool' or 'awesome') when first experiencing the VR presentation. When VR was the first condition that was experienced, the subsequent material was, in some cases, considered 'boring,' or it was stated that 'it was better in VR.' The time of exposure to the presentations varied between subjects, with a tendency to spend more time in the virtual environment, but the exact time for each condition was not measured. Several surgeons complained about the light reflection points on the surfaces and about the low resolution (independently of presentation). Two surgeons suggested adding parameters (e.g., an electrocardiogram and infusion pump information) to make the situation in the operating room more perceivable. All surgeons were able to intentionally examine the situs (e.g., zooming into specific structures), and no one reported any problems with the implemented controls or natural interaction in VR, respectively, which implies that the implemented interaction techniques did not negatively affect the ratings.

8.2.4 Discussion

The proposed MirrorMount for the HoloLens 2 has successfully been tested under real-world conditions and demonstrates the ability to use HoloLens 2 for recording and streaming data in telementoring applications. The analysis of the data of our study shows that the feeling of presence, as measured with SPES, is significantly increased, and surgeons feel more bpresent at a surgery from afar using a VR presentation in contrast to a 2D presentation. A 3D view on a screen can also increase the feeling of presence compared to a 2D video, but this effect is considerably weaker than that of a VR presentation. This

suggests that the process of telementoring might be improved to some extent by using a 3D reconstruction as expert surgeons from afar could feel more involved in the activity around the operational situs and the operating room. Furthermore, the evaluation of motion (MQ3) and heart function (MA2) was rated best in the VR presentation. The increased perception of motion and inherently better evaluation of cardiac function can probably be linked to the improved spatial perception in stereo vision using a head-mounted display. In contrast to SPES as a standardized tool to measure presence, the MQ and MA scales were provided by us to systematically capture the impression of surgeons. The high range of responses in each category may suggest that the questions were not precise enough or that subjective biases played an important role. Furthermore, the within-subject differences showed a high variation for some items (especially visible in fig. 8.12a, 8.12d, 8.13a, 8.13b and 8.13d), which also implies individual differences that need further investigation. MQ2 (p=.12), MQ3 (p=.14) and especially MA1 (P=.054) and MA4 (p=.076) were borderline not significant. With further improvements of the system and more elaborate studies, some findings may be made in the future regarding these aspects.

The results of the presented study are, however, quite limited as we only showed a short clip without much specific surgical content (such as anastomosis techniques or tissue assessment). The time spent on the presentation was considerably low, and deeper insight could have been obtained by presenting a more complex operation. Although the content seems a reasonable choice as a baseline, more advanced scenes that include the use of tools, surgical techniques, or specific moments within a surgery where a decision has to be made may be suited to analyze more sophisticated, specific research questions. Considering that five surgeons experienced VR for the first time, the ratings may also be slightly biased.

With the current system as presented in this work, various applications exist. While the system is not able to reconstruct fine details, which would be necessary for a precise diagnosis from afar, it can produce an immersive experience of medical procedures. Especially large structures, such as moving organs and tissue (for example, a beating heart), can be examined. The additional low-resolution point cloud enables an overview of the events in the operating room. It is also imaginable to use the obtained recordings as an immersive educational tool that visualizes operations in 3D. Students or doctors in training could use this to learn, e.g., about specific medical conditions, surgery techniques, and rare cases.

The reconstruction with data of a single depth sensor becomes much more difficult when hands or tools are involved, as parts of the situs are constantly occluded. Multiple sensors recording the situs from different angles would be suited to eliminate these movement-induced local occlusions to a high degree. For example, to further improve the coherence of the point cloud, the utilization of two or more HoloLens 2, as well as potentially stationary sensors (e.g., built into the operating room lights), which are registered in the same world space, seems promising. By combining the different views on the situs, both occlusions and missing pixels could be reduced, interpolated, or reconstructed. At this point, the quality of the depth sensor of a single HoloLens 2 seems barely sufficient to support the intended telementoring setting, as the resolution and noise do not allow for a precise reconstruction to make fine details visible. However, in the presented study, a work-in-progress prototype representing the state of the system in early 2022 was evaluated. During this stage of development, the transmission used a different method from what is presented in this thesis, resulting in a resolution of 960 $\times$ 540 pixels for the complete field of view, corresponding to approximately 300x200 pixels in the relevant situs area (see [Dew+22b] for details). This is a considerably lower resolution than what can be achieved using the final state of the prototype, which captures the situs area with a four times higher resolution of approximately 600 x 400 pixels. Nevertheless, the results are, to some degree, generalizable

for immersive telementoring systems with a similar resolution under real-world conditions. Whether a system with a higher resolution would have produced different results has not been investigated and remains a task for future research. In the future, we are especially interested in evaluating advanced records of surgeries that show actual procedures and finding ways for bidirectional communication that exploit 3D reconstruction, such as labels and drawn lines, as well as the transmission of video and possibly the inclusion of avatars. IVEs would also allow for multiuser applications that enable teams of surgeons at different locations to discuss a specific operation, which would be an interesting overarching topic for VR research regarding interaction techniques, communication, self-representation, and social factors in this field of work.

8.3 Summary

The presented study shows that the obtained 3D reconstruction from a single HoloLens 2 allows for an immersive experience of operations from afar in VR. Surgeons felt more able to evaluate cardiac function, which implies a possible usefulness in the medical domain. The Following key findings were made with our current prototype:

- In surgical applications, the camera view direction of the HoloLens 2 can be deflected to enable recording of the region of interest without forcing an artificial posture on users.
- The 5G network transmission of 3D data is possible using WebRTC, although fine details are lost due to encoding.
- A 3D reconstruction is possible using data from the built-in sensors of the HoloLens 2, although the resolution is rather low, and it is not clear if it will be sufficient for this use case.
- The HoloLens 2 can be worn above surgical loupes to provide an on-demand view of annotations.

The implemented mesh reconstruction from depth data and RGB video streams can be considered a classical approach for this task. In the future, with more advanced hardware, other methods may prove to be a better choice for performing this task. Recently developed candidates with high potential for producing high-quality images for immersive telementoring are, for example, Neural Radiance Fields (NeRFs) [Gao+22] and 3D Gaussian radiance splatting [Ker+23]. Currently, these approaches are, however, only aimed at reconstructing static scenes, not fast enough for real-time use, and they would require a sophisticated multi-camera setup, which cannot be provided in the intended use case of remote heart explantation, so using these approaches was not considered in this thesis. Overall, an application of these techniques in heart transplantation is not unrealistic in the near future, considering the fast progress in the area of artificial intelligence nowadays.

The limited field of view of the HoloLens 2, however, only makes the display of content in a small part of the user's field of view possible. Tests showed that it is possible to wear the HoloLens 2 on top of typical surgical loupes with individually fitted oculars. This suggests using the HoloLens 2 in this prototype system as an auxiliary monitor that displays 3D-registered annotations and other media content in the upper part of the field of view, while providing a clear view of the situs in the lower part of the field of view. By tilting their head, virtual content can also be aligned to the situs, however, from an ergonomic point of view, the usability of this head posture is questionable and has to be further investigated in the future. This type of operation has successfully been tested in a similar way with Google Glass [Yoo+17].

From the initially formulated goals **G1** and **G2** (see section 7.2.4), only the first goal was met using the presented approach. The prototype system is, from a technical perspective, suited for mobile use in the operating room during organ transplantation, and the real-time reconstruction in VR achieves a quality that is suited for an assessment of organs. However, the quality achievable using the HoloLens 2 as a single sensor does not achieve the millimeter accuracy required for the precise positioning of labels in a heart surgery context.

CHAPTER 9

OBJECT EXPLORATION

One of the main tasks of surgeons who remotely assess an organ for transplantation is determining whether an organ is suitable for both the transplantation procedure and the patient. In today's practice, surgeons evaluate hearts intended for transplantation by visually inspecting specific structures and movements of the heart and by performing manual examinations. The visual evaluation can easily be conducted in VR, however, due to the limitation regarding the visual display quality of many contemporary consumer devices (typically, a third or half the resolution of the human eye), the task can benefit from artificial interaction techniques that enhance the visual details. One approach to equalizing the low display resolution is the magnification of the details that are investigated. In this chapter, two super-natural exploration techniques suited for the intended use case and user group are investigated: Zoom Mechanisms and Hand-based Move-&-Scale. Both approaches are described on a technical level, and the former, as a novel interaction technique, is analyzed in a study.

9.1 Exploration Techniques

9.1.1 Zoom Mechanisms

In the real world, we are accustomed to magnification devices, for example, microscopes, binoculars, or reading glasses. In recent years, with the rise of mobile end devices and permanent access to the digital world, zooming into virtual content has become a common task in digital media consumption. Even though the 'pinch' gesture is a well-known way to interact with mobile devices to enhance the details of a 2D image, it has not yet been transferred to IVEs to enhance a 3D environment. In the prototype system for heart transplantation presented in this thesis, zoom mechanisms can be included as super-natural interaction techniques, as a simple way to enable surgeons to magnify organ structures (see Fig. 9.1) and fine details beyond the display resolution of the HMD, which is necessary to enable the application in real-world scenarios.

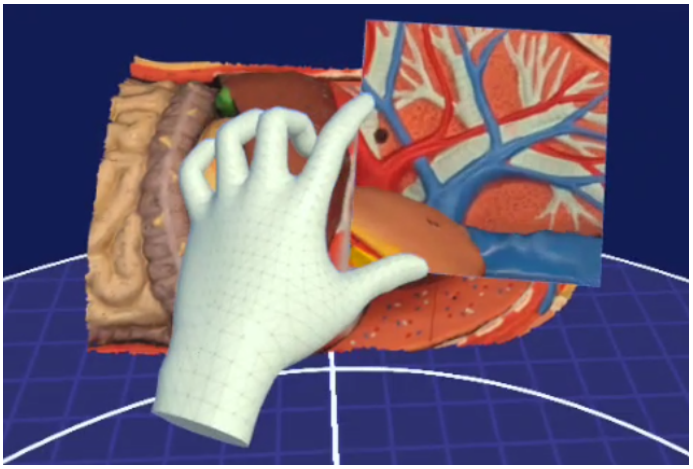


Figure 9.1: Quad lens used in the prototype system to inspect vessel structures.

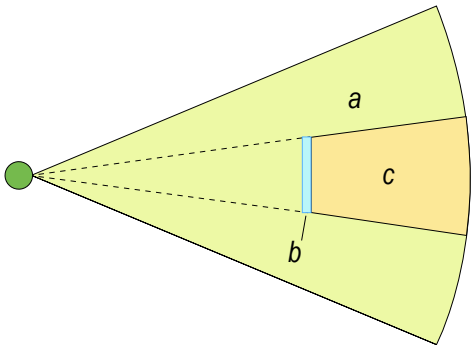
We developed two approaches based on finger tracking to perform an interactive dynamic zoom technique in IVEs: (i) *Aperture Zoom*, which mimics the operating principle of a real-world telephoto lens and is a more conservative way of displaying magnified content in 2D, and (ii) *Portal Zoom*, which maintains the 3D depth impression while zooming. Both approaches extend the magic lens metaphor [Bie+93], which allows for zooming into specific areas of interest to enhance the perception of details. This enables users to visually explore a large IVE from a stationary position ('telescope metaphor') and to magnify small details ('magnifying glass metaphor'), which can be difficult given the visual quality of current head-mounted displays. To evaluate the usability of both zoom mechanisms, we implemented both approaches and performed a user study to measure selection performance while zooming and to compare it with naïve ray casting as a baseline, and we captured the subjective impression of both interaction techniques.

Several zoom mechanisms in IVEs that enhance either the general details in high-detail panorama images [CC17; Mir+19] or specific details [Kna+19; CWLJ12] have been studied. Furthermore, the control of a real-world zoom camera coupled with a virtual reality camera has been investigated [KRR18]. In these cases, the control was based on the user's head movement [CC17; CWLJ12] or relied on an additional input device [Kna+19; Mir+19]. Zoom mechanisms have been visualized as a decal lens that is projected onto geometry [Mir+19] or as a modified small portion of the users' field of view [Kna+19; Gad+14; CC17]. Some research papers target a different form of zoom that locally scales 3D objects and can be related to the *World in Miniature* [SCP95] concept [Büs+19; WHB06]. Hand-held interaction concepts derived from the magic lens metaphor [Bie+93] in IVEs have been investigated (an overview of diverse implementations is provided by Tominski et al. [Tom+14]). Those concepts can be used to overlay additional information or provide an alternative visualization [Mar+19; LBC04]. Some of them have been adapted for target selection [Pie+97; Loo+07] and other forms of interaction with objects [SES99]. Although the application of zoom mechanisms was mentioned in [LBC04], an analysis of magnification mechanisms in terms of usability and task performance has not yet been conducted. Overall, while both hand-held magic lenses and zoom mechanisms have been studied, a combination of both, as presented in this chapter, has not yet been presented.

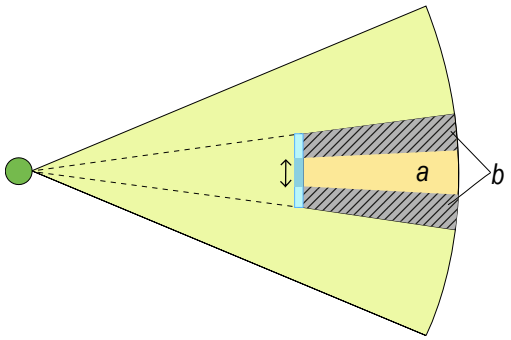
ApertureZoom

In a 2D environment, the effect of a magic lens can be achieved with an additional camera rendering pass that renders the area behind the lens with different rendering camera parameters (see Fig. 9.2b). For a zooming interface in 3D, the content behind the magic lens can be magnified by scaling the field of view of the rendering camera reciprocally to the intended magnification factor and displaying the content enlarged on the surface of the magic lens. This can, however, create occlusion areas behind the magic lens that are neither rendered in normal view nor in magnified view (see Fig. 9.2b).

Aperture Zoom works similarly to a magnification objective in the real world by decreasing the camera's field of view and thus magnifying the displayed objects reciprocally. Enlarging rendered content can be achieved by scaling the virtual camera's field of view to the inverted intended magnification factor and displaying the image on the surface of the magic lens. Without zooming, distinct volumes emerge behind the lens, which can cause binocular discrepancies and unpleasant effects such as mosaic rivalry when content is rendered once with a normal view for one eye and modified for the other (see Fig. 9.3a). Using a zoom mechanism, this problem is amplified as multiple conflicting volumes exist that render content differently. In the case of *Aperture Zoom*, only the dominant eye is rendered as a lens object as suggested by Forsberg et al. [FHZ96]. By this means, instead of creating a full 3D impression of a virtual scene, *Aperture Zoom* only renders a 2.5D image that



(a) A see-through lens. The frustum (a) is partially magnifyied (c) and projected onto the surface of the lens (b).



(b) A magnification lens. The frustum behind the lens is partially magnified (a) and scaled to fit onto the surface of the lens. Some areas behind the lens (b) are occluded and are not visible.

Figure 9.2: Illustration of the effects of a single lens on the field of view.

contains basic depth cues, such as perceived size, occlusion, and parallax effect similar to a real-world telescope. The zoom effect is implemented using an additional camera that points from the user's dominant eye towards the position of the virtual lens. The viewing angle of this additional secondary virtual camera is controlled by the intended magnification f_{AZ} using the relation $f_{\mathrm{AZ}} \cdot \alpha' = \alpha$.

PortalZoom

Portal Zoom, the second investigated technique, uses a translation $\vec{t}$ of an additional virtual camera along the view direction of the primary camera to achieve a magnification while maintaining the original FOV. The distance d' that is needed to obtain a magnification factor f_{PZ} without altering the overall FOV of a camera can be calculated using basic trigonometry (see Fig. 9.4). First, the visual angle of a target α with the visible size S at the distance d can be determined as:

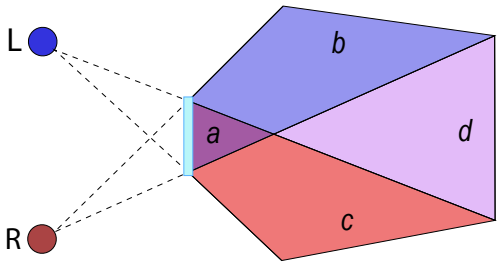
$$d \cdot tan\left(\frac{\alpha}{2}\right) = \frac{S}{2} \Leftrightarrow \alpha = 2 \cdot atan\left(\frac{S}{2d}\right) \tag{9.1}$$

Second, d' can be calculated using f_{PZ} as the intended magnification factor to modify the initial viewing angle:

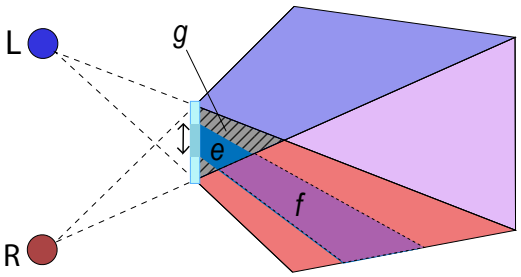
$$d' \cdot tan\left(f_{\mathrm{PZ}} \cdot \frac{\alpha}{2}\right) = d \cdot tan\left(\frac{\alpha}{2}\right)$$

$$\Leftrightarrow d' = d\frac{tan\left(\frac{\alpha}{2}\right)}{tan\left(f_{\mathrm{PZ}} \cdot \frac{\alpha}{2}\right)} = \frac{S}{2 \cdot tan\left(f_{\mathrm{PZ}} \cdot atan\left(\frac{S}{2d}\right)\right)} \tag{9.2}$$

As seen in equation (9.2), d' is dependent on both the size of the target S as well as the initial distance from the object to the camera d. Therefore, this zoom mechanism magnifies objects of the same size differently depending on the respective initial distance. Furthermore, a maximal zoom factor ${f_{\mathrm{PZ}}}^*$ exists. It is reached when d' is reduced to zero and can be calculated with $\frac{\alpha}{2} \cdot {f_{\mathrm{PZ}}}^* = 90°$.



(a) A stereoscopic see-through lens (L is left eye, R is right eye) without magnification. The area behind the lens is altered for both eyes (a). Some areas are affected by the lens for one eye while being normally visible with the other (b,c). The area far behind the lens is normally visible (d).



(b) A stereoscopic magnification lens as implemented in *Aperture Zoom* (zoom only for left eye). The magnification leads to occluded areas behind the lens (g), and only a small part is magnified without conflict (e). A conflicting area emerges that displays the content magnified for one eye and normal for the other (f).

Figure 9.3: Illustration of the effects of a virtual lens with stereoscopic vision.

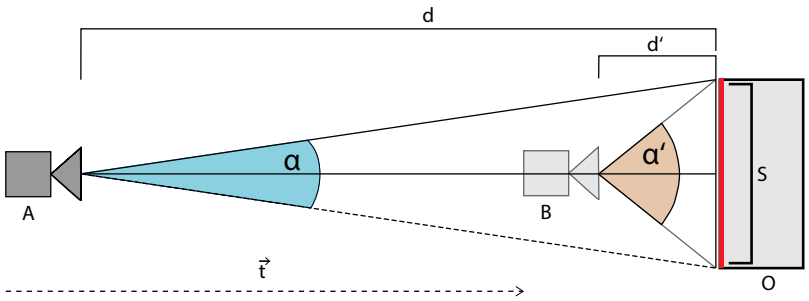


Figure 9.4: Illustration of *Portal Zoom*. The main stereo camera A renders an object O of size S from a distance d. A second stereo camera B is translated by $\vec{t}$ towards the object.

Portal Zoom uses an additional stereo camera and a pair of virtual lenses (one for each eye), which allows for a full 3D impression of the magnified content. The name 'portal' zoom derives from the fact that the lens acts like a window to an alternative visual perception of virtual content from a closer perspective. Technically, an additional stereo camera creates a complete second rendering, which is then superimposed onto the original rendering using the stencil buffer with the lens object as a mask. When the images from both lenses are merged in the brain through the convergence or divergence of the eyes, an artificial horopter is created. Points that are outside or in front of this horopter can be simultaneously perceived as two offset images – an effect called *diplopia*. *Portal Zoom* causes the horopter to shift in a non-natural way, which becomes visible during the transition phase from focusing on the hand in the foreground or the virtual objects in the background to the alternative perspective provided by *Portal Zoom* (see a depiction in Figure 9.5).

Both zoom mechanisms can be controlled using finger tracking. The lens is placed between the tips of the thumb and index finger of the dominant hand, and its orientation is aligned with the user's view. The distance between the thumb tip and index tip s_{TI} is continuously calculated and linearly mapped to the magnification factor and the size of the virtual lens. This leads to a mid-air pinch gesture, similar to the one used for zooming into digital content

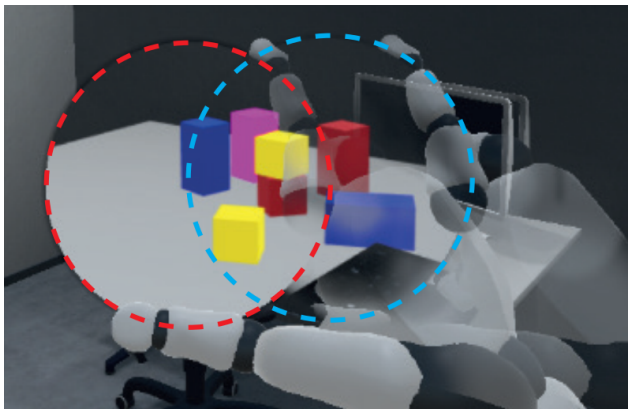


Figure 9.5: Visualization of perceived diplopia using *Portal Zoom*. Red: left eye, blue: right eye.

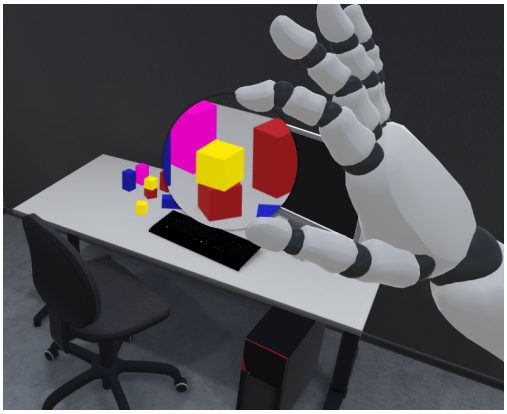


Figure 9.6: Aperture Zoom with 3x magnification.

in mobile-device interaction, which locally magnifies the virtual scene. The magnification factor f in our prototype system was calculated as $f = 5 \cdot (s_{\mathrm{TI}} - 1\ cm)/10\ cm$ and clamped to [1;5] to achieve a maximal magnification factor of 5. During the development, this value was determined as an adequate limit before involuntary body tremor interferes with aiming at specific locations too much, but still large enough to be suspected of having practical applications. The threshold of 1 cm was chosen to allow the visual alignment of the lens and an object without any magnification.

Additionally, a selection action was implemented for when an index-to-thumb tap is performed using the non-dominant hand. In the case of *Aperture Zoom*, this emits a single invisible ray from the lens's position along the vector between the lens and the dominant eye, which is tested for intersection with virtual objects in the IVE. *Portal Zoom* creates two individual rays that originate in the left and right cameras and pass through the position of the virtual lens. The average of the direction of both rays is set as the direction for a single selection ray cast that is emitted from the center of the virtual lens. In both cases, the hit point of the selection ray casts is visualized with a small magenta sphere at the hit location, with a diameter of 0.5 cm. All positional tracking data was filtered using a *one Euro filter* [CRV12].

9.1.2 Hand-based Move-&-Scale

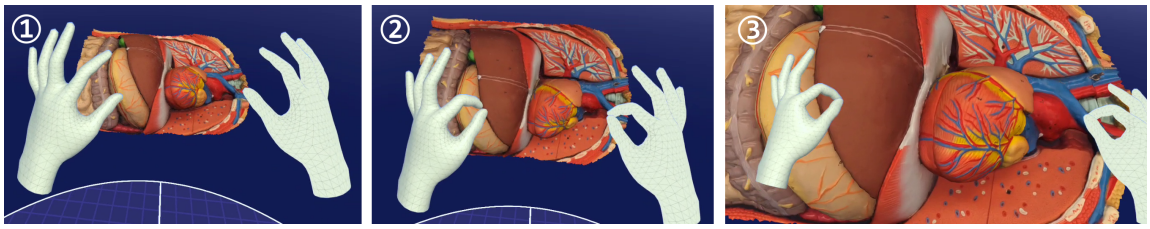


Figure 9.7: Hand-based Move-&-Scale. The bimanual pinch activation posture ①. Translating and rotating the virtual object ②. Scaling of the objects ③.

The second super-natural technique implemented is Hand-based Move-&-Scale. This technique is similar to the "Spindle" technique proposed by Mapes and Moshell [MM95] or the handlebar metaphor [Son+12], however, it utilizes the hands as input instead of a pair of controllers. To activate the interaction, the user performs a tap-&-hold posture of the thumb and index finger of each hand as a semaphor [KS05]. This can be interpreted as the reduction of the sensorimotor schema of grabbing an object to a similar, small-scale

movement. In this implementation, an object selection is skipped for the sake of simplicity. In the intended use case of HTX, only the situs is expected to be explored, making the indication and selection of an object obsolete. Holding a thumb-to-index pinch gesture synchronously with both hands activates the manipulation mode and places the situs in the middle between the user's hands. This reduces the number of interface-specific controls, properties, and visualizations, especially affordances, that have to be learned by users, for example, handles, which are commonly used for this task in 3D engines. The only gesture they have to internalize resembles holding and manipulating an object in reality.

To determine the position, rotation, and scaling factor, the vectors provided by Quest 2 built-in hand tracking model are exploited (see Fig. 9.8). The hand tracking yields 3D position vectors for the center of each hand as well as forward vectors describing the orientation of each hand ($\vec{f_R}$ and $\vec{f_L}$). From the positions of both hands, P_R and P_L, the center point P_C can be calculated as the average of both vectors, at which the manipulated object is placed, which enables the translation of the object. For the rotation, the object's right vector is calculated as the normalized vector from P_L to P_R with $r_O = norm(P_R - P_L)$. A temporary forward vector is calculated as the normalized average of $\vec{f_R}$ and $\vec{f_L}$ with $\vec{f} = norm(0.5 \cdot (P_R + P_L))$. A cross product between $\vec{f}$ and $\vec{r_O}$ yields the up vector $\vec{u_O}$, and a cross product of $\vec{r_O}$ and $\vec{u_O}$ yields the object's forward vector $\vec{f_O}$, fully describing the object's rotation. For scaling, the distance between P_L and P_R is calculated and mapped to a scaling factor.

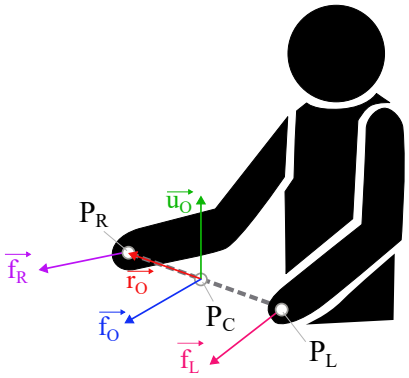


Figure 9.8: Diagram of the vectors used in the Hand-based Move-&-Scale technique.

Using this approach, the object can be moved similarly to how a real object would be manipulated using two hands by synchronized bimanual movements of both hands. While this may seem natural to the user, it is the result of carefully mapping the rotation of hands to the rotations of the object, rather than physically simulating a direct manipulation of virtual objects. Overall, this can be considered a schema adaptation, in which natural movements are mapped to similar movements of the manipulated object, effectively mimicking a direct manipulation of the object, even though, technically, this is a mapping of movements to the object's transformation. By increasing the distance between both hands, a scaling factor is determined that is applied to the situs. If the distance is below 0.15 cm, the situs maintains its original size. Otherwise, the distance maps exponentially to a uniform scaling of the object. This is an action that is hardly encountered in reality. Pulling a rubber band or a balloon can be related to this action. However, the exponential scaling and maintaining of the relative dimensions of the object differ from reality, making this a form of schema adaptation for users who are familiar with multi-touch pinch gestures that are typically used to magnify images on mobile devices. Otherwise, it can also be considered a form of sense-making in which the meaning of magnifying an object to explore small details is coupled to a distinct coordinated movement pattern of fingers.

9.2 Evaluation: Zoom Mechanisms

9.2.1 Motivation

To estimate the practical application of both zooming mechanisms, a formal exploratory study was conducted that investigated how easily users can zoom into a specific area of interest. This task is comparable to a selection task in which a specific object has to be indicated, but with the addition of a visual magnification. Therefore, the experiment task was designed as a task for indicating a virtual target in a Fitts' Law experiment while using the zoom mechanism with a certain magnification level, in our case, at least 80 % of the maximal zoom factor. The experiment was performed with both zoom mechanisms, *Aperture Zoom* (AZ) and *Portal Zoom* (PZ), and naïve ray casting (NR) without a zoom mechanism was used as a baseline selection technique in the same procedures to allow an estimation of speed and accuracy. In the experiment, we measured the amount of time needed to perform a selection (TIME), the error distance from the center point of target objects (ERROR), the error rate of missed targets (MISS) (when ERROR was larger than the radius of a target) and the effective throughput (TP) depending on selection technique (TECHNIQUE) using the calculations presented in [MI08]. Furthermore, the System Usability Scale (SUS) was used to rate the overall usability of both zoom mechanisms in IVEs.

9.2.2 Study

Hardware and Setup

A Samsung Odyssey was used as VR HMD in our study to display the IVE. One of the Odyssey's controllers was used for ray casting in condition NR. The finger tracking system utilized Optitrack cameras to reconstruct finger poses from infrared markers (see Fig. 9.10) with low latency (approximately 1 frame) and high accuracy (<2mm). Moreover, the HMD and the controllers were equipped with additional infrared reflective markers to align tracking spaces and stabilize positional tracking. A PC equipped with an Intel i9 processor and a GTX 2080 Ti was used to render the virtual scene at a constant frame rate of 90 Hz and to calculate the tracking data. The application was developed in Unity version 2019.2.9. The virtual scene setup shows a neutral office room in gray hues without any furniture in the participant's view (see Fig. 9.9). It was designed to provide a realistic impression of a neutral IVE without distraction.

Participants

We recruited 21 participants (15 male and 6 female), mostly students from our university. They were between 21 and 34 years old (M = 26.7, SD = 3.49), and all participants have worked with IVEs before. Moreover, all of them had a background in digital media design or computer science. They had normal or corrected-to-normal vision; color vision was tested using color charts, and depth perception was tested using random dot stereograms. The Participants were allowed to take a break or completely terminate the experiment at any time, but none of them wished to do so.

Material and Methods

As a first step in the experiments, the dominant eye was determined using the Dolman method ('hole-in-the-card') test, and the experiment was prepared according to the result. Next, we tested the ability to see color and depth. The respective participant was then instructed to stand on a marked position in the tracking space and not leave it in order to

maintain a constant distance from the targets of the Fitts' Law experiment. As the last step in the preparation, the subject received the HMD and hand-tracking gloves, both of which were then calibrated (interpupillary distance and hand segment lengths). During the preparation, the nature of the experiment was explained, and participants were instructed on how to use the three different interaction techniques in our experiment. Afterward, five single targets were shown to the participants to try out all selection techniques (see Fig. 9.10) until they felt confident applying them.

Spheres with a diameter of 20 cm (visible angle of 3.81°) were arranged in a circle with a diameter of 1 m. Their color changed according to their current status in the experiment: *neutral* (blue), *target* (red), *selected* (transparent dark blue). The center point C of the Fitts' Law reciprocal task circle was placed at a height of 1.5 m and a distance of $d = 3$ m to the user's ground-projected head position P during the experiment (see Fig. 9.9). The distance between the circle and the participant's HMD was spatially constrained to retain a constant distance. The background of the Fitts' Law circle did not provide any visual cues, so target selection was primarily based on the aim of a participant.

Procedure

The highlighting of current targets started with the one at the center top and continued with the opposing target in a counterclockwise motion. We chose to use 11 targets in the experiment as a tradeoff between the number of repetitions and avoiding fatigue of the participants. Moreover, the participants were instructed to be 'as fast and precise as possible' when aiming at a sphere and to avoid missing a target. The error distance ERROR was determined by projecting the hit point of a selection from the surface of the sphere onto the plane that passes through the sphere's center and faces the participant. ERROR is, therefore, the distance between the projected point and the center of the sphere. A target is missed when the distance is greater than the sphere's radius. During the experiment, the data points were automatically logged in our application and later analyzed in SPSS. To minimize carry-over effects, we randomized the order of interaction techniques. After the formal experiment, the participants were given the chance to try out the zoom techniques and explore the virtual scene freely. During this period, they were instructed to state their thoughts ('thinking aloud' method), which were noted down. The experiment concluded with a final question regarding which zoom technique was preferred by the participant. Finally, they filled out the SUS and NASA TLX questionnaires for AZ and PZ.

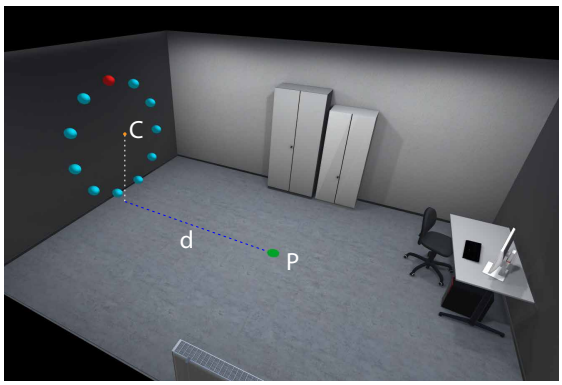


Figure 9.9: A neutral office environment with no distractions as IVE.

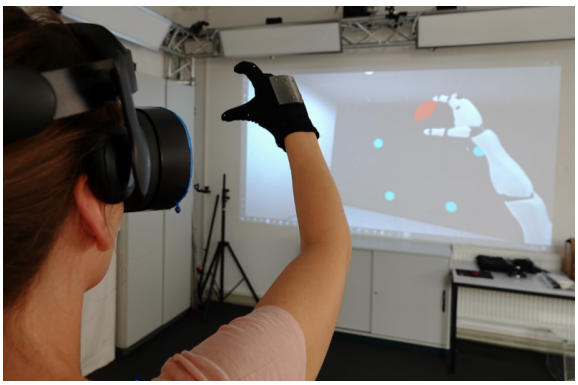
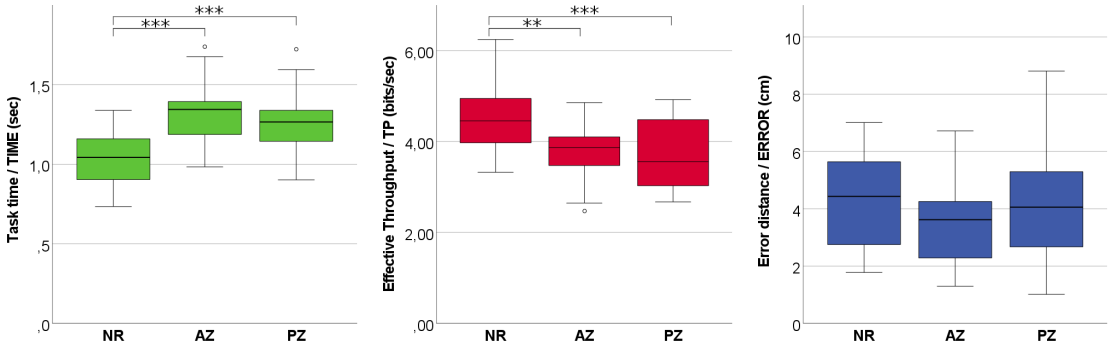


Figure 9.10: Pinch gesture used for target selection during the experiment.



(a) Median task time. (b) Effective throughput. (c) Error distance.

Figure 9.11: Visualization of the data distribution of different conditions as box plots.

9.2.3 Results

The measurements of the experiment are presented in Table 9.1 and further visualized in Fig. 9.11. Table 9.1 displays mean and standard deviation for normally distributed data as well as median, lower (Q1 = 25 %), and upper (Q3 = 75 %) percentiles for non-normally distributed data. To calculate TP, we followed the procedure presented in [Mac18]. The results of Kolmogorov-Smirnov test were not significant for TIME, TP, and ERROR, so a normal distribution of data was assumed. Additionally, the results of Levene test were also not significant for TIME ($F(2,60) = .038, p = .963$), TP ($F(2,60) = .529, p = .592$), and ERROR ($F(2,60) = .926, p = .402$). Therefore, a one-way repeated-measure ANOVA was conducted with post-hoc repeated-measure pairwise comparisons at 5 % significance level (with Bonferroni correction). The results of Mauchly test did not show significance for TIME ($\chi^2(2) = 2.566, p = .278$), TP ($\chi^2(2) = 5.790, p = .056$), and ERROR ($\chi^2(2) = 3.576, p = .166$), so a correction of degrees of freedom was not required. In the case of MISS, only some participants missed one or two single targets, and the Kolmogorov-Smirnov test showed significance ($p. < 001$), so the experiment was analyzed using the Friedman test for repeated measures. The results of the Kolmogorov-Smirnov test were also significant for SUS scores, so the Wilcoxon signed-rank test was used for a comparison of AZ and PZ.

Dependent Variable	Unit		NR	AZ	PZ
Median Task Time (TIME)	s	*Mean (SD)*	1.04 (0.18)	1.33 (0.20)	1.26 (0.12)
Effective Throughput (TP)	bits/s	*Mean (SD)*	4.45 (0.75)	3.76 (0.65)	3.69 (0.43)
Error Distance (ERROR)	mm	*Mean (SD)*	42.3 (3.6)	35.1 (2.9)	41.9 (4.8)
Error Rate (MISS)	ratio	*(Q1, Median, Q2)*	(0,0,0)	(0,0,0)	(0,0,0)
System Usability Scale	points	*(Q1, Median, Q2)*	-	(92.5,97.5,97.5)	(92.5,95,97.5)

Table 9.1: Measurements for median task time (TIME), Effective Throughput (TP), error distance (ERROR), error rate (Miss) und SUS points.

The effect of TECHNIQUE on TIME was significant ($F(2,40) = 19.041, p < .001, \eta^2 = .488$). Post-hoc tests revealed a significant difference for NR vs. AZ ($t(20) = -5.450, p < .001$) and NR vs. PZ ($t(20) = -5.674, p < .001$) but not for AZ vs. PZ ($t(20) = 1.215, p = .7152$). TECHNIQUE had a significant effect on TP ($F(2,40) = 12.226, p < .001, \eta^2 = .379$). Post-hoc tests revealed a significant difference for NR vs. AZ ($t(20) = 3.346, p < .01$) and for NR vs. PZ ($t(20) = 4.627, p < .001$) but not for AZ vs. PZ ($t(20) = .536, p = 1$). The effect of TECHNIQUE on ERROR was not significant ($F(2,40) = 2.097, p = .136, \eta^2 = .085$). In the case of MISS, the Friedman test was not significant ($\chi^2(2, 21) = 3.250, p = .197$). The results of the Wilcoxon test did not show a significant difference in the SUS ratings ($Z = 23.5, p = .378$).

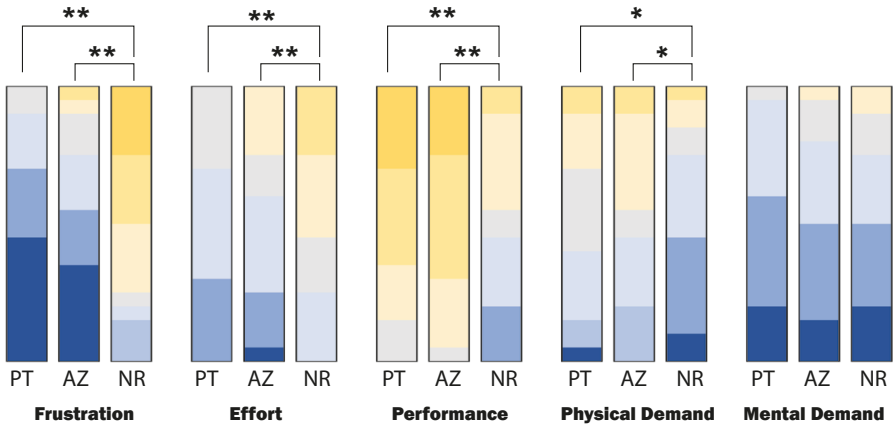


Figure 9.12: Distribution of NASA TLX answers (blue: low, yellow: high).

Subjective Rating

The distribution of subjective ratings in E1 is displayed in Figure 9.12 and analyzed using the non-parametric Friedman test. The only item that did not show significance was *mental demand* ($\chi^2(2) = 0.6944, p = .7066$). Physical demand ($\chi^2(2) = 11.1944, p < .01$), performance ($\chi^2(2) = 22.75, p < .001$), effort ($\chi^2(2) = 10.7778, p < .01$) and frustration ($\chi^2(2) = 23.5278, p < .001$) differed significantly. Post-hoc comparisons using the Wilcoxon signed-rank test revealed that there was no significant effect between AZ and PZ zoom for all items. AZ was rated significantly higher than NR for physical demand ($z = -2.511, p < .05$), performance($z = -3.724, p < 0.01$), and significantly lower for effort ($z = -2.580, p < .01$) and frustration ($z = -3.7893, p < .01$). PZ was rated significantly higher than NR for physical demand ($z = -2.417, p < .05$), performance ($z = -3.823, p < 0.01$), and lower for effort ($z = -3.29, p < .01$) and frustration ($z = -3.823, p < .01$).

Additional observations

When further asked about their impression of the zoom techniques, most participants described them as feeling 'natural' or 'intuitive' and compared the interaction to the pinch gestures typically used on mobile devices. Using their hand to control the zoom techniques provided 'some anchoring' of the virtual lens in the 3D scene, which facilitated the selection task.Some participants They especially liked the 'responsiveness' of the lens to the motions of their fingers. Two main strategies for zooming into a location emerged: Most participants first disabled the magnification to aim at a location and then magnified the content in a second step. Some of them maintained a constantly high zoom factor while moving the virtual lens to a location. Both strategies seemed to perform equally well. When asked which zoom technique they preferred, 14 of the 21 participants chose *Portal Zoom*. The main reasons for this were an improved 3D depth perception, an enhanced immersive feeling ('being just in front of the objects'), and more fun while exploring. Three participants commented negatively on the perceived merged lens shape and the comparably small area of depth perception while using *Portal Zoom* (see Fig. 9.5 for a visualization). Two had some trouble merging the two 2D images in the beginning, but were able to do so within a short amount of time. Four participants who preferred *Aperture Zoom* liked the feeling of having 'more control' ('cleaner interaction'). However, some participants did not like keeping their non-dominant eye closed while using AZ as it felt 'tiring' to do this for a prolonged period of time. Three participants liked both zoom mechanisms equally well. No participant reported symptoms of cybersickness.

9.2.4 Discussion

The results of our study show that both zoom mechanisms presented in this work can easily be applied by users in a virtual environment. Compared to naïve ray casting regarding the effective throughput and time, they perform slightly worse in simple selection tasks. Considering the intended use as a visual enhancement tool, however, the difference is negligible, as a straightforward interaction for deliberately selecting and magnifying areas of interest is possible with an appropriate speed and accuracy. Furthermore, it is important to note that participants neither felt disoriented nor reported an increased feeling of cybersickness, even though the visual impression of the IVE was highly manipulated. The results of the formal experiment, the SUS questionnaire, and the observations are coherent. Participants enjoyed the interaction and were visibly excited to use both zooming mechanisms for scene exploration. Even though *Aperture Zoom* and *Portal Zoom* differ concerning the technical implementation and visual effect, their performance is not significantly different. Moreover, both zoom techniques are easy to implement and can be utilized in many applications, given that the system is capable of performing additional renderings of the virtual scene. The interaction can be controlled using finger tracking, but the mapping of controls is easily transferable to other input devices, such as generic controllers, which makes a generic application in virtual environments possible.

Overall, we found that both zoom techniques, *Aperture Zoom* and *Portal Zoom*, do not differ significantly. An adaptation can be useful as a universal exploration tool for IVEs that enables users to explore the surroundings from a standing position or to magnify small details, for example, reading text or, in the intended use case in heart transplantation, examining small vessels and anatomical structures. The small difference to naïve ray casting selection considering task time and effective throughput implies a sufficient speed in applications. Both zoom techniques can easily be implemented in many applications to enhance the user experience and extend the catalog of interaction tools.

9.3 Summary

Both techniques, hand-based Move-&-Scale and the proposed zoom mechanisms (ApertureZoom and PortalZoom), are simple to learn and use, and all users who tested the interaction were almost instantly able to explore a virtual 3D object with high proficiency. As the implemented movements are purely physical and, from the user's perspective, no symbolic processing is involved in the interaction, the interaction can be considered entirely based on enacting interaction schemata that are highly internalizable. The congruence of the involved pattern can be considered low, considering that magnification is not typically experienced outside of VR applications. Both interaction techniques support the prototype's goal **G3**, the natural and effective interaction using only hands as input.

In the case of a VR reconstruction of an HTX surgery, the technique enables users to perform novel ways of exploring the situs that are not possible in reality, which effectively changes the post-phenomenological second-order relation to the real world if an unlimited camera resolution is assumed. In such a case, the VR application becomes a hermeneutical tool that can be used to explore the world, similar to a telescope. Regarding the first-order relation, both interaction techniques enable users to enhance details that are not easily visible in the IVE, which corresponds to an increased visual acuity among users. For example, reading small texts, which is still a problematic application due to the limited resolution of current devices, is facilitated. This form of improvement is also valuable in an immersive telementoring application. For example, if medical records such as X-ray images are integrated into a virtual environment, zoom mechanisms provide a quick and intuitive way for examination.

CHAPTER 10

SURFACE ANNOTATION

The second main task in the intended telementoring use case is the annotation of the virtual representation of the organ as 3D object. With a mobile HMD and only hands as input devices, as intended in the proposed prototype system, this task can be challenging. In contemporary devices, the camera-based tracking of hands is fast enough for real-time interaction, however, the accuracy and precision are lower than for controller-based interaction because hands are often primarily intended for gestural interaction in today's VR interface design. This makes the annotation of structures in the use case of HTX challenging, considering that the position of annotations, in many cases, needs to be highly accurate within millimeter precision. In this chapter, three super-natural techniques for hand-based annotation, Scale-&-Draw, LensDraw, and PalmDraw, are described in detail, and they are analytically evaluated [LJ+17] by analyzing the enaction of the techniques based on the implemented schemata and their mapping to functions (as presented in chapter 5.1.4). Furthermore, they are compared to a natural approach to annotation in VR in an empirical pre-study to show their potential in future research and applications.

10.1 Annotations Techniques

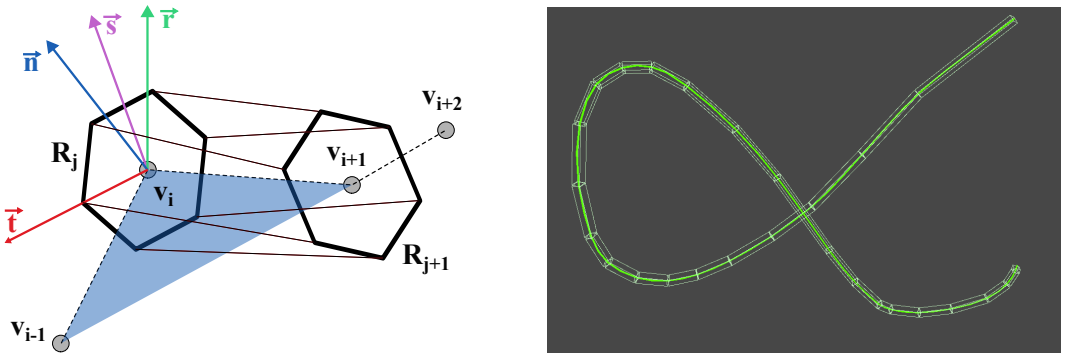
10.1.1 Astronaut's Tool Belt & Virtual Pencil

The Astronaut's Tool Belt is a support interaction technique for tool selection in VR that is used in the later-presented annotation techniques. The tool belt metaphor [For+00] is a well-known approach for tool selection in IVEs. In typical implementations, pressing a specific button opens a menu from which an item that corresponds to an assigned function can be selected. However, in the IVE created for HTX as the use case in hospitals, we intentionally wanted to include no controller and focused instead on whole-hand input to increase acceptance among surgeons. Without a controller as an input device, different ways would exist to open a tool belt, such as gestures and voice commands. However, this requires the user to learn a specific command that is mapped to the system-control function to open the menu, which is not always intuitive. Especially in large applications, where diverse commands are necessary to activate and control all functions, memorizing all commands and recalling them perfectly after a long time can be demanding. In the Astronaut's Tool Belt, the functions used in an application are represented by objects floating in front of the user, similar to objects floating in space without gravity. However, instead of floating freely in an environment that is not affected by gravity, the object is pulled toward a point in space at a fixed offset to the user's head.

In this annotation use case, the tool belt only holds one item: a pencil that naturally provides the affordance to be used by users for drawing. The pencil's gravity center is located approximately 30 cm in front and 20 cm above eye level, and, depending on handedness, 20 cm to the left or to the right. This enables a user to simply look up to locate a tool corresponding to a function, and by naturally reaching for it, the corresponding function is activated. Grabbing the pencil activates the drawing mode, in which the pencil is positioned in a natural pose in the user's hand. While we only use annotation as an interaction mode in our application, it would be easy to add a small number of different tools to the

Astronaut's Tool Belt that are ready to hand in an IVE, which is sufficient to support many simple applications. The Astronaut's Tool Belt combines super-natural aspects and natural interaction to form a simple-to-learn technique for tool selection. The floating of objects with some kind of stable gravity well in relation to the user's head is a behavior that cannot be experienced in reality (low experiential congruence). The selection of a tool can be considered an enhancement of being in an IVE that changes the user-environment relation and enables new interactions. Furthermore, the process of selecting a tool by identifying its location and reaching for it is highly internalizable and aligned with how we intuitively handle tools in reality. After several repetitions, it becomes even possible to internalize the position in relation to the head, facilitating the selection of a tool without requiring the user to look at the object first. When users finish interacting with the tool, they can simply release the finger posture, and the pencil floats back to its designated position.

The pencil used in the Astronaut's Tool Belt behaves similarly to a real pencil. As soon as the pencil's tip collides with an object marked for annotation in the IVE, a drawing function is initiated, which is executed as long as the pencil's tip stays on the object's surface. When contact with the object for annotation is lost, the drawing is terminated. No haptic feedback can be generated using this approach based on whole-hand input without specialized equipment, for example, in-air ultrasonic devices [Rak+20]. Using the classification system for schema analysis (see section 5.1.4), this can be interpreted as a *meaning altering* in which touching the virtual object with the pencil does not lead to a haptic sensation that would be evaluated by the user to control movement. This requires some enaction by users, which is, however, achieved after several seconds of interacting. To provide some feedback to users, the pencil tip plays an audio loop of a pencil drawing on paper when it touches a drawable surface to give the user some feedback for the drawing event, which can also be experienced as an auditory feedback during drawing in reality (*schema integration*).



(a) Construction of the volumetric line generated by the geometry shader.

(b) Procedurally generated box colliders for a volumetric line.

Figure 10.1: Implementation of the volumetric line generated by the pencil tool.

A geometry shader was implemented to render annotations in VR, as the lines generated by Unity3D's built-in LineDrawer exhibited instability and visual artifacts at this small millimeter scale, making it an unsuitable option for accurate annotation in the context of heart surgery. The shader constructs a volumetric line from a list of vertices that is constantly updated with points from the drawing event (see Fig. 10.1a). For each vertex $\vec{v}$, a triangle with the normal vector $\vec{n}$ is constructed from the predecessor and the successor (blue triangle). With a global reference vector $\vec{r}$, an orthogonal coordinate system $(\vec{t}, \vec{s})$ is calculated that is used to determine the vertex positions on a ring R with a radius corresponding to half the diameter of the line. Usually, the global up vector $(0, 1, 0)$ of the virtual environment is used as the reference vector $\vec{r}$, however, when the

determinant of $\vec{r}$ and $\vec{n}$ equals zero, the slightly rotated vector $(0, 0.995, 0.1)$ is used. The vertices of the subsequent rings are connected to create the new faces of the volumetric line. To enable users to interact with the line, for example, deleting or translating lines, procedurally generated box colliders are added to the volumetric line (see Fig. 10.1b). The implementation uses a greedy algorithm that iterates over the vertices of the line and generates boxes when the angle between the direction of a subsequent line segment and the initial direction of the box collider exceeds a defined threshold. The colliders are three times as wide as the drawn line to approximate the shape of the line while still being easy to hit. The line transformation component is set as a child of the virtual object's transformation component to enable synchronous transformation of the object and attached annotations.

10.1.2 Scale-N-Draw

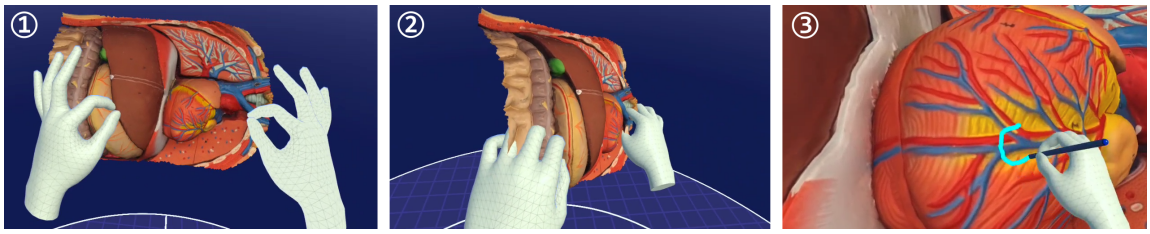


Figure 10.2: Drawing surface annotations directly on the object in 3D.

Scale-N-Draw combines hand-based Move-&-Scale and the pencil tool in Astronaut's Tool Belt into a technique for annotation (see Fig. 10.2). Using the whole-hand-based Move-&-Scale technique (①), users are enabled to freely move and uniformly scale (up to a scaling factor of 5) an object for annotation in the environment. After adjusting the position in a way that is beneficial for annotation (②), the pencil tool can be selected from the Astronaut's Tool Belt to activate the drawing mode. Annotations are performed by directly drawing on the virtual object in 3D (③), which is an efficient technique for annotation [Yu+22]. Users can easily change their perspective on the object by releasing the pencil, which then floats back to its location in the tool belt, adjusting the position, rotation, and scale of the object, and then grab the pencil from the tool belt again.

Function	**Schema**	**Type**	**Coupling Mechanism**
Activate object transformation	Bimanual pinch gesture	Semaphoric finger posture	Sense-Making
Object translation / rotation	Synchronized bimanual movements	Direct manipulation (mapping)	Schema adaptation
Object scaling	Synchronized bimanual movements	Mapped hand gesture	Schema adaptation
Switch to drawing mode	Grab the virtual pencil	Semaphoric finger posture, metonymy	Schema integration
Draw on virtual object	Move pencil on the object's surface	Direct manipulation	Schema integration

Table 10.1: Schema analysis for the most important interaction schemata in the Scale-&-Draw technique.

The embedded schemata (see Tab. 10.1) follow a reality-based approach without replicating realistic interaction. To initiate object interaction, a pinch gesture is performed simultaneously with both hands, which can be classified as a semaphoric finger posture. This gesture can be freely executed in space to bring the object instantaneously towards the user. While it resembles the act of grasping an object in reality, its novelty suggests it can be better understood as a form of sense-making. Similarly, object transformation is performed through synchronized hand movements that resemble the hand-based manipulation of a real object. However, it is not a direct manipulation but rather a mapped movement that is applied to the object's translation and rotation. Annotation, on the other hand, is carried out by drawing directly on the surface, which can be viewed as schema integration, as this action mirrors how annotation could occur in the real world.

10.1.3 LensDraw

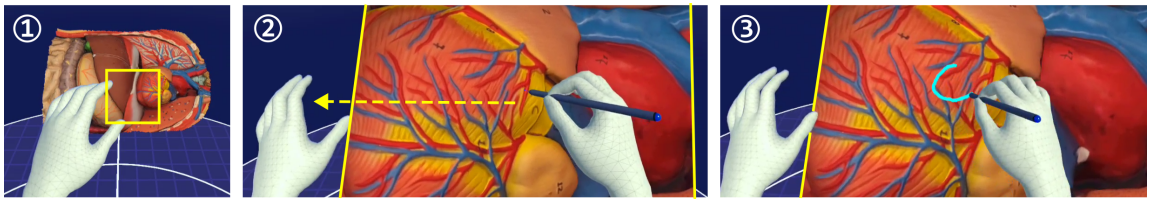


Figure 10.3: Drawing surface annotations indirectly on a scalable magnification lens using a virtual pencil. The yellow border and arrow are not present in the IVE and were added to enhance the comprehensibility.

Similar to Scale-&-Draw, LensDraw combines the previously presented LensZoom technique (see chapter 9.1.1) and the pencil tool in Astronaut's Tool Belt into a compound technique (see Fig. 10.3). After grasping the pencil from the Astronaut's Tool Belt, in addition to placing the pencil in the user's dominant hand, a magnification lens is placed between the thumb and the finger of the user's non-dominant hand (①), which is used as a planar projection of the object for indirectly annotating structures. In this implementation, the lens has no magnification factor for a distance smaller than 5 cm between the thumb's tip and the index finger's tip. The magnification factor reaches its maximum (5x) at a distance of 10 cm with a linear interpolation between 5 and 10 cm (a mapping finger gesture). This enables users to easily locate an area of interest and adjust the magnification factor rapidly. When the pencil is hovering above the lens, the position and orientation of the lens are locked, and increasing the distance between the lens and the non-dominant hand increases the size of the lens to a diameter of up to 50 cm (②). With the magnified view of the virtual object, the user can directly draw on the planar lens object (③), and annotations are projected onto the object behind the lens.

In this technique, the integrated schemata (see Tab. 10.2) can be considered less reality-based. The use of a virtual object that is strongly coupled to hand movements resembles real-world tools but exceeds the typically expected effects of physical objects. There are similar devices in reality, for example, a smartphone camera with a zoom function, however, a direct analogy does not exist, therefore, the applied schema represents a form of sense-making. Similarly, the activation of the drawing mode and the extended magnification of the zoom lens in the second step of the interaction have no real-world equivalent. Annotation is performed by directly drawing on the magnified zoom lens, which can primarily be considered a form of schema integration from the user's perspective during interaction. However, projecting these annotations onto the virtual 3D object introduces an additional layer of sense-making that requires some enacting.

Function	Schema	Type	Coupling Mechanism
Activate magnification lens	Grab pencil (pinch)	Semaphoric finger posture, metonymy	Sense-Making
Increase lens magnification	Move index and thumb	Mapped finger gesture	Sense-Making
Prepare drawing	Hover pencil above lens	Semaphoric finger posture	Sense-Making
Increase the size of the locked lens	Move non-dominant hand away while hovering the pencil above the lens	Mapped hand gesture	Sense-Making
Draw on a virtual object	Move the pencil on the object's surface	Direct manipulation	Schema integration

Table 10.2: Schema analysis for the most important interaction schemata in the LensDraw technique.

10.1.4 PalmDraw

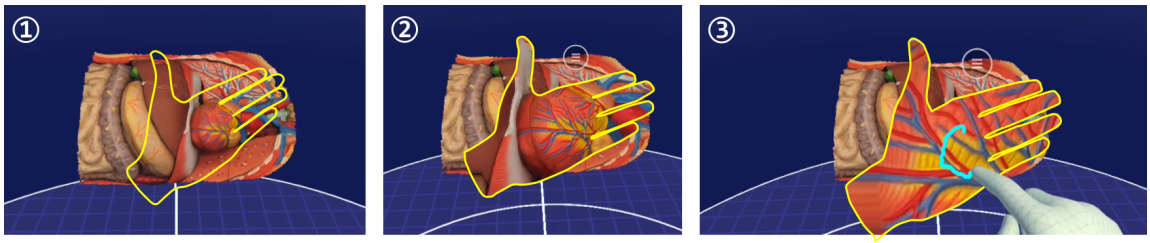


Figure 10.4: Proxy drawing on the non-dominant left hand using the index finger of the dominant right hand. The yellow borders are added to the figure to enhance differentiation between the hand and the background, which is not necessary in the virtual scene.

PalmDraw uses the hand as a target for indirect annotation as a see-through object and the dominant hand's index fingertip to perform annotations (see Fig. 10.4). ① The users can see the model for annotation through their non-dominant hand. By changing the location of the hand, users are able to aim at a specific location by looking through their hand. ② When the hand is moved towards the user's head, the view is continuously magnified up to a factor of five. ③ Moving the index fingertip of the dominant hand on the palm of the non-dominant hand creates annotation lines that are projected onto the model.

Technically, this is implemented using an additional virtual camera at the head position of the user that is aimed at the position of the user's non-dominant hand, rendering the scene into a 512 x 512 RenderTexture. Using a shader, this texture is projected in screen space onto the geometry of the hand. The hand object is masked from the rendering layer of the second camera, allowing the user to effectively 'see' through their hand. When the palm faces the user (a semaphoric hand posture), the rendering is activated, otherwise, the hand is rendered as a virtual hand model. Moving the hand toward the user's head is mapped to a decrease in the field of view, resulting in a zoom function (mapping of a moving posture). When a user enters a bounding volume with their dominant hand, the current transformation of the non-dominant hand is stored as T_{hand}, and the position and orientation of the additional camera are locked to provide a static view. To draw annotations on the geometry projected onto the hand, the user uses the index finger to

draw shapes on the projected rendering with the non-dominant hand as a proxy surface that offers a haptic sensation. To detect drawing events, the vertices defined by Meta Quest 2's hand model of the hand are constantly retrieved, and a low-poly mesh is modified according to the world position of 13 vertices (see Fig. 10.5). This low-poly mesh is used in Unity to calculate a MeshCollider that detects collisions with the index finger when the index fingertip enters the MeshCollider's volume. Using the inverse transformation of the current hand and the stored transformation T_{hand}, the collision position can be transformed from world space into the initial locked hand tracking space to perform a ray cast that draws an annotation at the intersection with the reference model.

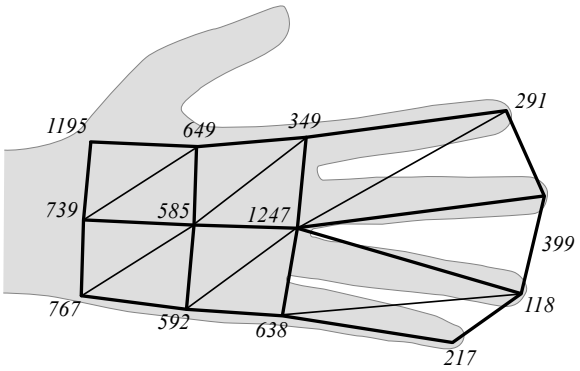


Figure 10.5: Mesh topology and exploited vertices of Meta's Quest 2 hand model used for the low-resolution mesh collider.

The interaction schemata implemented in this technique (see Table 10.3) differ from reality-based interaction. Physical actions, such as turning the hand, are reinterpreted in a novel way to provide new functionality, which makes this form of annotation mostly a form of sense-making that has to be enacted. In contrast, drawing on the palm can be seen as a form of schema adaptation, in which the act of painting on a surface is approximated. Similar to LensDraw, this technique introduces a sense-making layer when the annotations performed on the hand are projected onto the virtual 3D object.

Function	**Schema**	**Type**	**Coupling Mechanism**
Activate annotation mode	Palm faces upwards	Semaphoric hand posture, metonymy	Sense-Making
Increase magnification	Move hand closer / further away	Mapped hand gesture	Sense-Making
Prepare drawing	Hover index finger above palm	Semaphoric finger posture	Sense-Making
Draw on hand	Draw onto palm using index finger	Direct Manipulation	Schema Adaptation
Draw indirectly on mesh	Draw onto palm using index finger	Direct Manipulation	Sense-Making

Table 10.3: Schema analysis for the most important interaction schemata in the LensDraw technique.

10.2 Pre-Study: Annotation

10.2.1 Motivation

In this section, the presented annotation techniques are analyzed to compare how well they are suited to perform annotations in the setting of heart transplantation. The use of super-natural interaction techniques requires a careful design of interaction. In contrast to transferring interaction from reality, the design space is considerably large, and not all approaches work as intended, making incremental improvements necessary after the initial development. In contrast to other types of user studies in HCI, which often measure fixed properties such as reaction times or cognitive load, the evaluation of experimental interaction techniques may also focus on aspects that developers can directly influence. Poor performance may not necessarily stem from a bad approach but often from a flawed or unoptimized implementation. Additionally, the hardware used has a significant impact on the results and potentially contributes to performance issues or limits the technique's effectiveness. This makes performing a pre-study an appropriate step during development, in which design flaws or false assumptions are captured early on and considered in subsequent development steps.

Especially in the presented approach that utilizes whole-hand input as the primary means of interaction, the design of the interaction can have a large effect on the annotation performance. Research regarding whole-hand input shows that this type of input usually generates more cognitive load during tasks [GDA19] and typically has a lower performance in comparison to other means of input [RMC15; Ran+23]. In the intended use case of heart transplantation, this is a challenging finding, considering that this application requires a fast and accurate way of interaction. The speed-accuracy tradeoff [MI08] is often a challenging aspect, and identifying approaches that are both sufficiently fast and accurate is not a simple task. Therefore, in the following section, a study is presented that analyzes the previously presented interaction techniques for annotation in terms of their annotation speed and accuracy. Their common theme is providing internalizable ways of lowering the control-display ratio to increase annotation accuracy.

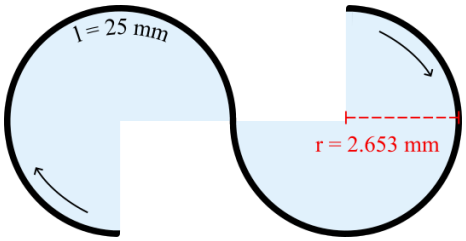
10.2.2 Study

Hardware and Setup The study application was developed in Unity 2022.1.20f1 and deployed to a Meta Quest 2 VR HMD to enable conducting the experiment using mobile VR devices. All steps in the study were carried out in VR in a single application, and the experiment was conducted at different locations.

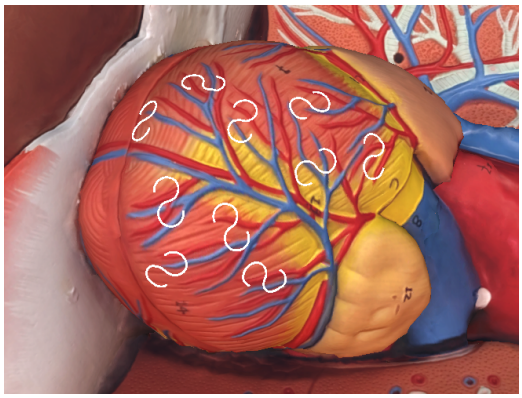
Participants For this preliminary study, we recruited 7 participants (age: mean=28.3, SD=2.3). Six were male, one was female. All had a background in computer science or related fields and had some experience in VR. One was left-handed, and all others were right-handed.

Material and Methods In this pre-study, four annotation techniques were analyzed: virtual pencil as naïve implementation and baseline technique, LensDraw, Scale-N-Draw, and PalmDraw. The photogrammetry scan of artificial organs of the upper human torso was used as the object for annotation to demonstrate a use case as close to reality as possible. A two-circle figure (see Fig. 10.6a) with a diameter of 10.6 mm was used as the annotation target line. This shape was selected as it provides an inflection point during the annotation, which is assumed to require more control than a simple line or circle. The size of the figure corresponds to annotations that could be used to mark specific vessels or structures in the use case of heart transplantation. The position of annotations is randomly

distributed across the 3D model of the heart (see Fig. 10.6b). The size of the figure in each position was determined by projecting it onto the 3D model from a specific point and, using 200 sample points distributed along the figure, scaling it until the length of all edges between points approximated 25 mm.



(a) The two-circle figure used in the tracing trials.



(b) The 10 positions for line tracing projected onto a 3D model of the heart.

Figure 10.6: The figure used in the experiment as a tracing shape.

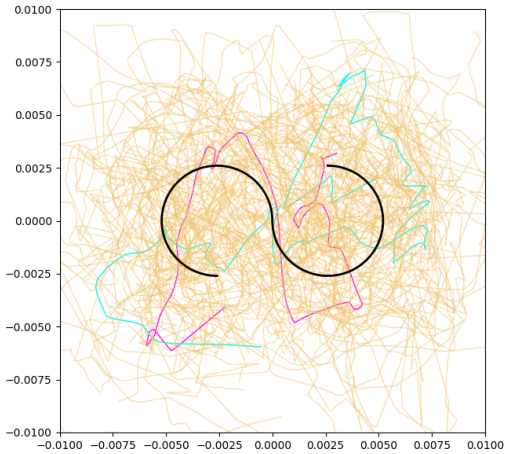
The application deployed on a Meta Quest 2 captures all annotation data automatically on the device during the experiment. For each trial, the application logs a list of vectors that describe the drawn annotation line, the timespan between displaying a target figure and the user starting to draw the annotation, and the timespan between starting the annotation and finishing the tracing. These were retrieved after the experiment and converted to data frames in Python. A well-known algorithm to measure the distance between two lines, in this case, a target figure and an annotation sample, is the Fréchet distance, which is calculated as the minimal maximal distance between two lines if both are traversed in parallel [EM94]. Using the Python package "similaritymeasures 1.1.0," the Frechet distance for each sample can be calculated. For the statistical analysis, the Python package "statsmodel 0.13.5" as well as the software PSPP were used.

Procedure The experiment began with an introduction to the study, in which participants were briefed on the basic idea of the study and familiarized with the Meta Quest 2. The participants were instructed that their primary priority was accuracy, and speed was only secondary, however, one annotation should not take longer than "3 or 4" seconds. Each participant experienced the four annotation techniques in a random order. At the start of each session, a video was shown that demonstrated the interaction technique. Participants were then given as much time as needed to familiarize themselves with each technique before they started the actual experiment. During the experiment, 10 figures were projected one by one in random order onto a virtual heart (see Fig. 10.6a), and participants were asked to trace each figure. Once a figure was completed, the next one was automatically projected until all 10 figures had been traced. After completing all figures, the subsequent interaction technique began automatically, and this process repeated until all techniques had been tested. An observer was present throughout the experiment to address any participant questions and to take notes on observations and remarks. Following the experiment, a debriefing session was conducted, during which participants were encouraged to revisit the techniques and freely share their impressions and experiences.

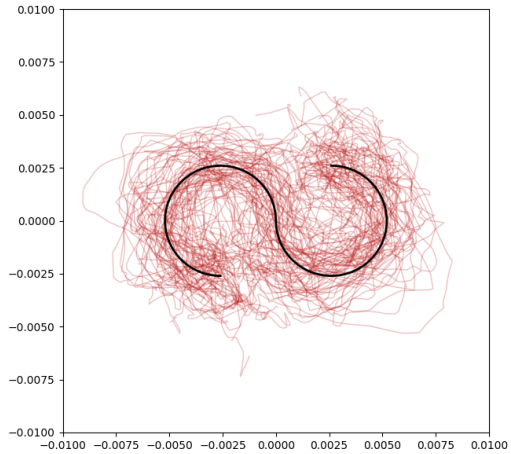
10.2.3 Results

Annotation Accuracy

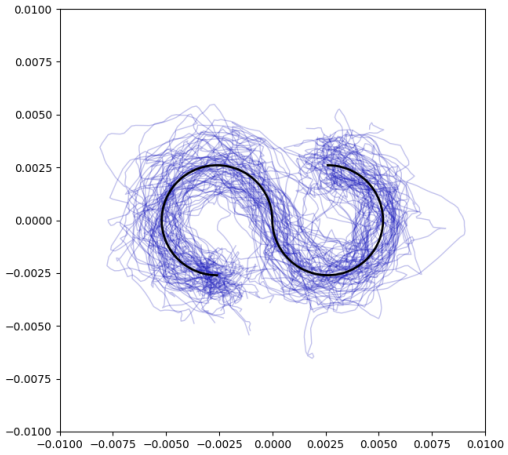
The line tracings for all techniques and participants are plotted (see Fig. 10.7). It is clearly visible that the virtual pencil technique performs worst in this scenario, and the resulting tracings (see Fig. 10.7a) do not approximate the intended shape. Two random samples are highlighted (cyan and magenta colors) to make the tracing performance in this case more comprehensible. It is visible that line tracings are indeed resembling the intended shape, but fail to reproduce it adequately. In contrast, the tracings of PalmDraw, LensDraw, and Scale-N-Draw all approximate the shape quite well, with Scale-N-Draw providing the best approximation, and LensDraw and PalmDraw showing still acceptable quality.



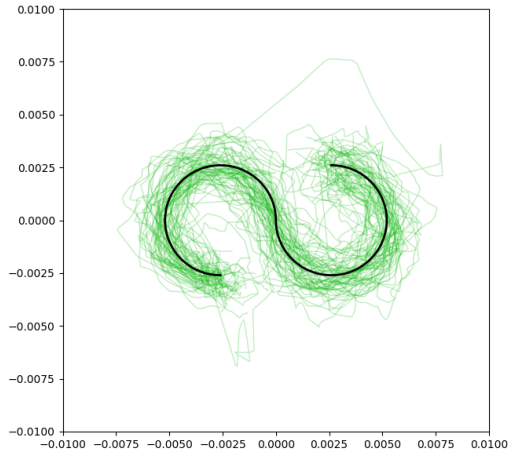
(a) Pencil (two samples highlighted).



(b) PalmDraw.



(c) LensDraw.



(d) Scale-N-Draw.

Figure 10.7: Plots of all line tracing samples of the four investigated techniques with the two-circle figure (black) as reference.

For each sample, a Fréchet distance (fd) is calculated. As an average measurement of the line tracing performance, the median of the fd for each technique is calculated per subject (10 sample curves). The mean of the median values is both resistant to extreme values and,

in this case, approximately normally distributed (Shapiro-Wilk tests were not significant). The mean of medians (fd_{mm}) per subject and technique is then calculated as an estimated average performance per technique. As subjects were allowed to begin tracing the shape from either endpoint, the fd is calculated twice, in the original sequence of vertex points and in reversed order, and the lower value of both is selected. The results are shown as violin and box plots in Fig. 10.8. In coherence with the path tracing visualization Fig. 10.7, it is visible that the virtual pencil technique performs worst (median: 5.78), Scale-N-Draw best (median: 1.77), and both LensDraw (median: 2.13) and PalmDraw (median: 2.72) approximate the reference shape slightly worse than Scale-N-Draw.

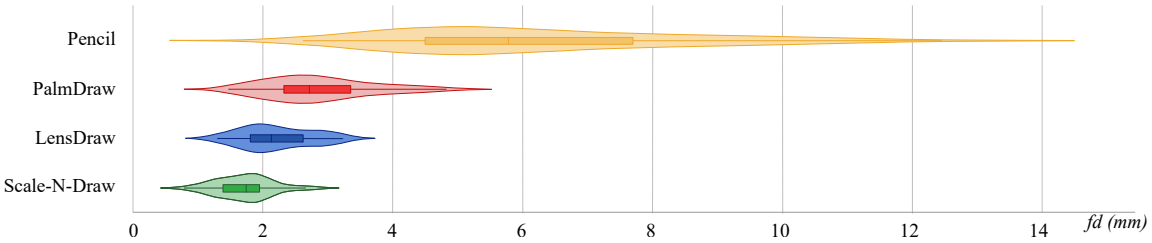


Figure 10.8: Violin and box plots of the measured Fréchet distances of all samples for all techniques. Outliers were removed using Tukey's fences.

In the case of fd_{mm}, Mauchly's test of sphericity indicated that the assumption of sphericity had been violated (Mauchly's test: $\chi^2(5) = 22.611, p = .001$), thus a Greenhouse-Geisser correction was used ($\epsilon = .385$). The repeated-measure ANOVA indicated a significant difference between techniques ($F(1.15, 18) = 42.801, p < .0002$) with a large effect size (partial $\eta^2 = 0.877$). Post-hoc Bonferroni-corrected ($\alpha = 0.0083$) paired t-tests for techniques were significant for most of the pairs except LensDraw and Scale-N-Draw, and the statistics confirm the visual impression obtained from plotting the tracings (see Fig. 10.7). Scale-N-Draw performs best ($fd_{mm} = 1.788, SD = .231$), followed by LensDraw ($fd_{mm} = 2.141, SD = .201$) and PalmDraw ($fd_{mm} = 2.756, SD = .374$), and the least accurate annotation technique is the pencil technique ($fd_{mm} = 6.120, SD = 1.53$).

	LensDraw	PalmDraw	Pencil	Scale-N-Draw
fd mean	2.212	2.859	6.262	1.731
fd median	2.131	2.721	5.782	1.768
fd_{mm}	2.141	2.756	6.120	1.788
SD (fd_{mm})	.201	.374	1.53	.231

Table 10.4: Descriptive statistics of the measured Frechet distances (fd and fd_{mm}).

Comparison (f_{mm})	difference	t statistic	p-value	Cohen's d
LensDraw vs. PalmDraw	.615	3.189	.019	1.21
LensDraw vs. Scale-N-Draw	-.353	-4.227	.006*	1.60
LensDraw vs. Pencil	3.980	6.894	.0005**	2.61
PalmDraw vs. Scale-N-Draw	-.967	-4.774	0.003*	1.80
PalmDraw vs. Pencil	3.364	5.175	.002*	1.96
Scale-N-Draw vs. Pencil	4.332	8.407	.0002**	3.18

Table 10.5: Pairwise post-hoc paired t-test comparisons fd_{mm}.

Annotation Times

In the same way as fd_{mm}, the mean median overall annotation time t_{mm} can be calculated from the raw drawing time t. In the case of t_{mm}, Mauchly's test of sphericity indicated that the assumption of sphericity had been violated (Mauchly's test: $\chi^2(5) = 11.501, p = .042$), thus a Greenhouse-Geisser correction was used ($\epsilon = .552$). The repeated-measure ANOVA indicated a significant difference between techniques ($F(1.66, 18) = 29.276, p < .0002$) with a large effect size (partial $\eta^2 = 0.830$). Post-hoc Bonferroni-corrected ($\alpha = 0.0083$) pair-wise comparisons for techniques were significant for some of the pairs, making the pencil technique the fastest annotation technique ($t_{mm} = 1.81$), followed by LensDraw ($t_{mm} = 2.08$) and Scale-N-Draw ($t_{mm} = 2.32$), and PalmDraw ($t_{mm} = 2.49$) as the slowest technique (see Tab. 10.7).

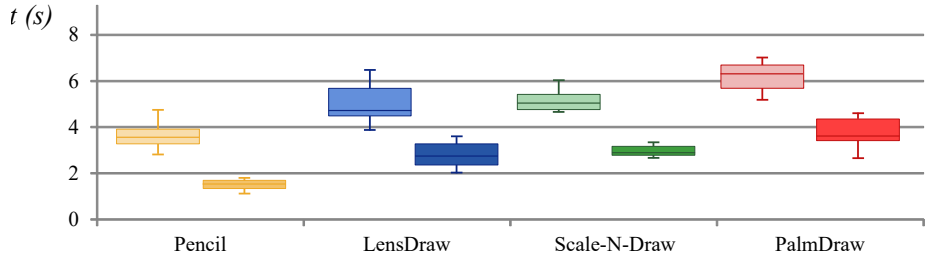


Figure 10.9: Boxplots of the overall annotation time t (shaded in lighter color) and preparation time t' (shaded in darker color) for each technique.

	LensDraw	PalmDraw	Pencil	Scale-N-Draw
t' mean	2.868	3.748	1.810	2.935
t' median	3.115	3.925	1.977	3.190
t'_{mm}	2.836	3.805	1.777	2.946
SD (t'_{mm})	.581	.587	.397	.255
t mean	5.361	6.428	3.797	5.480
t median	5.036	6.304	3.700	5.151
t_{mm}	5.112	6.301	3.710	5.239
SD (t_{mm})	.941	.684	.631	.548

Table 10.6: Descriptive statistics of the measured drawing times (t and t_{mm}) and preparation times (t' and t'_{mm}).

Comparison (t_{mm})	difference	t statistic	p-value	Cohen's d
LensDraw vs. PalmDraw	1.189	3.0175	0.0235	1.14
LensDraw vs. Scale-N-Draw	.127	.385	.7136	0.15
LensDraw vs. Pencil	-1.402	-8.183	.0002*	3.09
PalmDraw vs. Scale-N-Draw	-1.062	-4.230	0.0055*	1.6
PalmDraw vs. Pencil	-2.591	-9.5035	<.0001**	3.59
Scale-N-Draw vs. Pencil	-1.529	-8.486	<.0002*	3.21

Table 10.7: Pairwise post-hoc paired t-test comparisons for the annotation time of investigated techniques (N=7).

Subjective Aspects

Overall, the objectively measured results were supported by subjective feedback from participants. All participants found the pencil baseline technique very difficult to use and unsuitable for the intended scenario. This was primarily due to high noise levels, which led to inaccurate annotations. There was no clear consensus on which of the other techniques was preferred. All three proposed supernatural interaction techniques were quick and easy to learn, with some participants finding the Scale-N-Draw technique the most user-friendly. Both LensDraw and PalmDraw were also quickly mastered and became efficient after a few repetitions. Two participants highlighted the positive aspect of haptic feedback in PalmDraw, whereas others found it slightly more awkward to use than LensDraw and Scale-N-Draw. Furthermore, one female participant with smaller hands noted the limited surface area of the palm as a drawback.

10.2.4 Discussion

The low number ($N = 7$) of participants only enables drawing a tentative conclusion, which can provide hints about the analyzed effects. However, this still provides some insight that can influence the design of future studies, especially considering that the statistical analysis revealed highly significant differences between groups. The results indicate that all three approaches to supernatural annotation techniques can be effectively utilized in immersive telementoring and offer advantages over a basic implementation, specifically the virtual pencil technique. In terms of accuracy and speed, Scale-N-Draw proved to be a highly effective method for conducting annotations, while PalmDraw and LensDraw performed only slightly worse. However, since all three techniques were still in an early experimental phase at the time of evaluation, it is not yet possible to definitively determine which technique is best suited. Due to the various design possibilities for super-natural techniques, there are numerous factors that need to be determined in user studies. In the case of super-natural techniques, this definition often occurs arbitrarily in the initial stages of development, as there is no real-world reference technique that is replicated. This design space means that evaluations cannot be conducted in a straightforward manner but should instead be carried out incrementally and iteratively. Such an approach enables the correction of any false initial assumptions. Furthermore, and for PalmDraw in particular, the influence of the hardware must be considered, as the optical tracking system is challenged by hand overlap. Future generations of hardware may improve tracking quality and thereby potentially reduce negative hardware-induced effects on these techniques.

In this study, no difference was observed regarding the impact of the employed schema coupling mechanisms. Two major groups of interaction schemata can be distinguished: reality-based schemata, which incorporate or modify interaction patterns from the real world, and novel interaction schemata, which rely entirely on sense-making. The internalizability of interaction schemata appears to be independent of these groups, which suggests that there may be an 'intrinsic' difficulty in coupling schemata and functionality within an application, particularly for novel schemata. Since no difference in the duration of the coupling process was identified in this experiment, it can be hypothesized that the intrinsic difficulty of coupling new schemata may not necessarily be greater than that of transferring already familiar interaction patterns. The experiment further indicated that usability and internalizability are not the same. Although all techniques were quick to learn, they did not all perform equally in terms of practical usability. The baseline virtual pen interaction technique was unsuitable for annotations at the intended scale of less than one centimeter. Notably, opinions on PalmDraw were divided, and no clear consensus emerged from the study. This suggests that while a technique may be easy to internalize, it may still present challenges in terms of usability, particularly in tasks requiring fine detail.

10.3 Summary

All investigated annotation techniques combine super-natural elements (object scaling, a virtual magnification lens, a see-through hand, or a floating pencil) with natural means of interaction (mapping between hand movements and object translation, direct annotation in 3D) that are derived from real-world interaction. The preliminary study indicates that the interaction techniques can be considered super-natural, as they are easy to learn and apply, perform well in the designated task, and utilize principles that are not known from the real world. Implementing these types of interaction techniques can be beneficial in the context of immersive telementoring as they reduce both the preparation time of hardware, as no complex tracking system or controllers are required, and cognitive load during the interaction. At the current point in research, it can be assumed that the prototype goal **G3** can be supported using this approach.

A basic implementation of displaying annotations on the explantation side has also been integrated into the prototype, which is an important aspect of a complete prototype system. This can be achieved relatively easily as the 3D data is available in spatial relationship to the HoloLens 2, which enables the overlay of virtual elements onto the real environment. However, a challenge for future development is that the surgical wound is not static but instead constantly moving, or it may be shifted or obscured by the surgeon. To address this, suitable methods must be developed that track the tissue and enable a stable and precise attachment of annotations to specific structures. AI-based methods present a promising solution and can be explored in future projects to make the prototype usable in real-world scenarios. Beyond heart transplantation, the techniques presented offer a potential for a variety of applications. PalmDraw and LensDraw, for example, can not only be used to mark small details on surgical wounds but can also enable remote annotations on any virtual object. These techniques could also be applied in AR applications to quickly attach virtual annotations to objects in the real world.

PART IV

OUTCOME & REFLECTION

CHAPTER 11

DISCUSSION

11.1 On Enactivism in VR and HCI

The enactive approach offers a rich and valuable framework for analyzing human-computer interaction in the context of VR. It provides the foundation for a conceptual framework developed in this thesis, which can be summarized as follows: For many applications in diverse domains, certain aspects of IVEs can be considered to constitute an artificially created avatar-virtuality system, which is enacted by the user as a cognitive agent in the process of sense-making. On the one hand, this system can provide novel quasi-physical and conceptual properties in the user's virtual surroundings as novel affordances. These artificial affordances can be directly perceived by the agent in a similar way to how real-world affordances are perceived [WS20]. To the user, as an embodied cognitive agent, the rules that describe how the real world is enacted also apply to the exploration and mastery of a virtual environment; In the moment of enaction, virtuality seems real. On the other hand, the user may embody [Zie03] an avatar with an altered virtual body through which he or she enacts meaningful ways of structural coupling to the environment. Using this body to interact with the environment follows the same principles as using the real body in a real environment. Although the user's virtual body and the virtual environment are artificially constructed, the ability to interact with an environment is not inherently dependent on recreating reality. Instead, as long as the configuration of the agent-environment system makes sense, enaction will occur, and system-specific cognition will emerge as a foundation of acting and behavior in an IVE.

Many types of interaction in VR, especially interactions that primarily involve physical actions rather than abstract operations, can be considered a type of System 1 cognition [Eva03] which primarily relies on constructing and retrieving schemata for interaction [Swe03]. Interaction schemata can be described as a hierarchical network consisting of conceptual ideas about interaction and sensorimotor schemata (normatively evaluated closed-loop systems that have a particular function [DPBB17]) that describe the physical components of interaction. Distinct interaction schemata can be analyzed in terms of their internalizability, which determines how easily a schema can be integrated into the user's system cognition. Both the avatar and the environment form a developer-determined co-determination that was intentionally created by a VR developer and offers certain agent-world relations [Ihd17] (both 1st order and 2nd order). Often, this co-determination is designed to enable actions that are beneficial to the agent and activity in which the agent is present, sometimes purposefully to solve specific tasks, and sometimes for the sake of experimentation or entertainment. From interaction in virtual environments, IVE-specific short-term patterns of self [Gal13] emerge. These emerging patterns of self characterize the type of entity that exists as an agent in a specific environment. Users can experience diverse patterns of self as congruent or non-congruent to their typical lived experience, depending on how easily and naturally they can adopt a specific pattern of self. This mostly depends on the existing long-term patterns acquired in prolonged and repeated embodied interaction in other agent-environment systems.

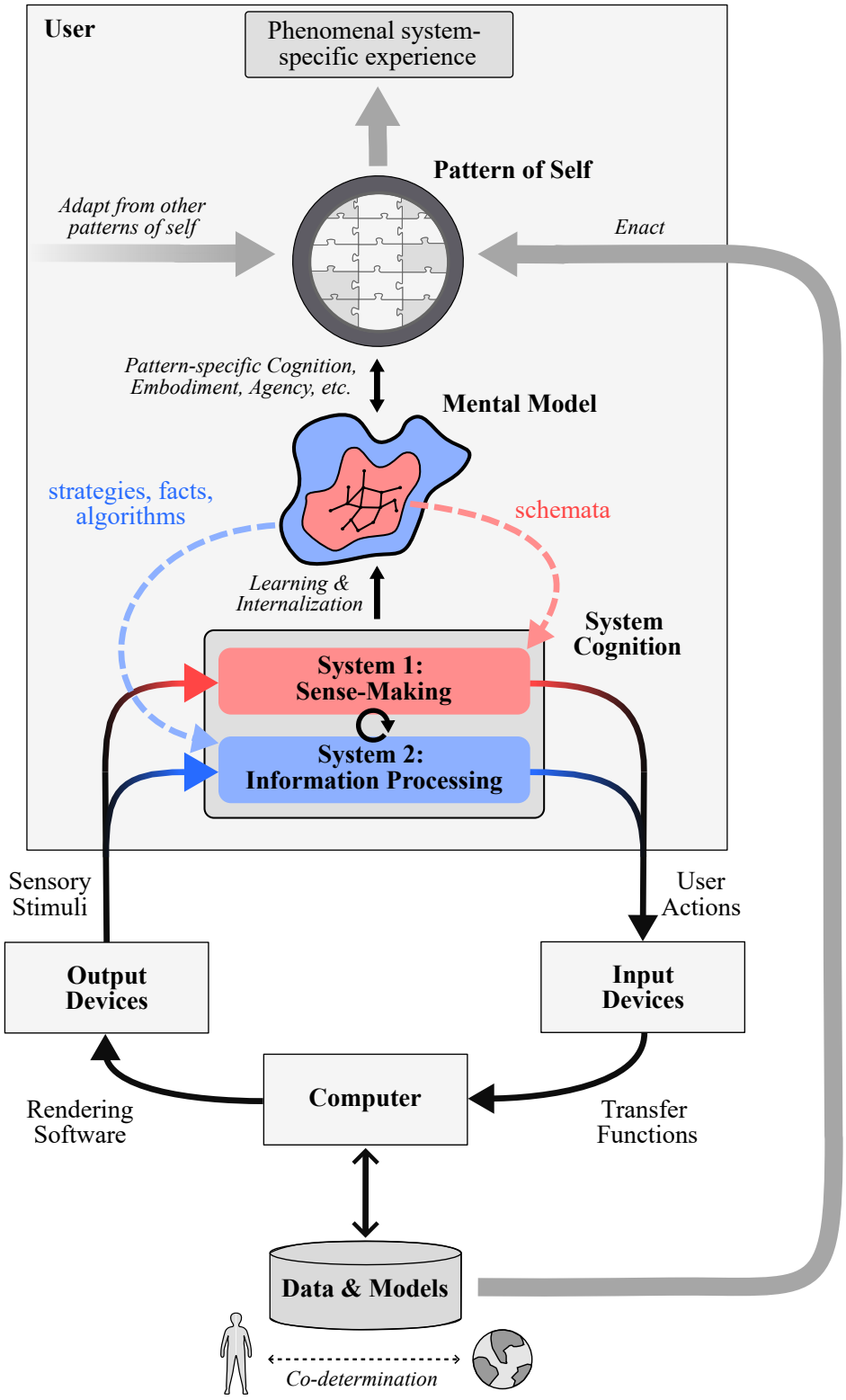


Figure 11.1: System 1 and 2 (integrated into the FIFA framework [MLP16]) as intertwined modes of system cognition during user-computer interaction. The mental model is integrated into the enacted pattern of self and provides schemata and strategies for interaction.

Although this conceptual view on interaction in IVE enables new perspectives on HCI in ICEs, which are usually not considered in contemporary scientific discussions, many of these insights have to be considered preliminary. Constructing a truly exhaustive, unifying framework can be challenging in complex and multifaceted fields such as HCI. The approach presented in this thesis focuses on the phenomenological and individual perspectives on interaction by incorporating ideas of enactivism and related philosophical concepts. Other important aspects of contemporary HCI, such as the social and cultural dimension, are only considered at the periphery and, depending on the application, require a much deeper analysis. However, by integrating the developed conceptual research into a practical application, the enactivism-based approach seems fruitful, and exploiting or further refining the presented concepts seems a worthwhile endeavor. For the interaction between a single user and a computer system, especially VR systems, the FIFA model by McMahan [MLP16] can be expanded (see Figure 11.1) to include enactive aspects. Since data and models in a virtual environment do not have to follow the rules of the real world, the co-determination of an IVE becomes an interesting aspect for future research. These aspects may complement interaction frameworks regarding subjective and antirealist effects.

The physical interaction and the construction of interaction schemata, which combine conceptual knowledge with physical action in the form of sensorimotor schemata, are essential aspects in this regard. We believe that one strength of VR lies in its character as a highly enactive medium [Kai+11], in which sensorimotor loops and System 1 cognition based on internalized schemata are of great importance for interaction. This emphasizes considering the role of the physicality of interaction when users interact with VR systems. In traditional computer systems, the physical aspect of input and output is typically limited to pressing mouse buttons and keystrokes, as well as a limited amount of related fundamental operations [FWC84]. In contrast, in VR, there are numerous possibilities for designing user interaction and a much richer catalog of possible actions. This means that the physicality of interaction, which plays only a secondary role in traditional human-computer interaction, is, in VR, not necessarily focused on enabling abstract symbolic input and output but instead a core aspect of immediate and embodied interaction. Furthermore, this makes providing diverse user input and output channels an important aspect, as both are required for closed-loop systems, in both realistic and antirealistic interactions.

The presented research regarding enactivism in VR was mainly conducted from an interpretative perspective [GP90] that directs the focus of research toward subjective effects. This aligns well with themes of third wave [Bød15] or paradigm [HTS07] HCI, in which it is argued that functionalist [GP90] approaches cannot fully capture the complexity of interaction with computer systems. Gioia and Pitre describe that "the goal of theory building in the interpretive paradigm is to generate descriptions, insights, and explanations of events so that the system of interpretations and meaning, and the structuring and organizing processes, are revealed" [GP90], which is a fundamentally different approach compared to the dominant functionalist paradigm in HCI that "seeks to examine regularities and relationships that lead to generalizations and (ideally) universal principles" [GP90]. However, both can complement research in multiparadigm perspectives on research and theory building [GP90]. A true combination of interpretative and functionalist approaches could not be fully achieved in this thesis, primarily due to the absence of quantitative methods within the framework of enactivism and post-phenomenology. Therefore, this thesis is limited to i) providing a framework for research through enactivism and promoting approaches derived from enactivist considerations and ii) quantitatively investigating the resulting effects from these approaches.

11.2 On Super-Natural Interaction

The first part of this thesis investigates the term 'super-natural' interaction. The lack of definitions and conflicting descriptions was identified as a shortcoming in the current research. The term 'super-natural' is inconsistently used in VR research. Often, it serves as an adjective describing interaction design approaches rather than representing a clearly defined concept. Although many researchers share intuitive ideas about its meaning, interpretations vary, and a universally accepted definition has yet to be established.

To address this problem, a conceptual model has been developed based on the enactive approach and related philosophical concepts. The pattern theory of self, post-phenomenology, and the concept of organ projection offer rich frameworks and underlying considerations that support the construction of the conceptual ideas presented in this thesis. The enactive framework offers valuable insights into super-natural interaction by focusing on the user's embodiment in VR and the co-determination of the user-environment system. This aligns well with the physical nature of VR interactions and the creative freedom in designing immersive virtual environments that differ from reality. The derived dimensions, internalizability, congruence, and enhancement, support the application of super-natural interaction techniques and provide a complementary perspective to other human-computer interaction frameworks, such as information processing theory. Super-natural interaction techniques are defined by their ease of use and learning (high internalizability), their departure from real-world constraints (low congruence), and their ability to alter the user-world relationship in a beneficial way (high enhancement). These three dimensions form a conceptual model called the ICE cube, which spatially relates super-natural techniques to other interaction design approaches.

To some extent, this thesis introduces its own terminology to provide definitions. Even though *congruence* and *internalizability / internalization* are no new terms and have been used to express similar concepts in other contexts, the use as presented in this thesis is not common in contemporary HCI research. In some cases, using more established terms could simplify the terminology of the presented models and reduce the required conceptual knowledge presented in the sections 4.1 and 5. For example, instead of using 'congruence' as a dimension in the ICE cube, 'realism' or 'perceived realism' would presumably lead to similar results. Likewise, 'learnability' would, in many cases, be rated similarly to 'internalizability.' In both cases, however, the introduction of new terminology allows for a precise definition and a clear reference to related concepts. After all, learnability and realism can be interpreted in different ways, and research shows that, in both cases, it is often not fully clear what exactly is referenced when researchers employ these terms (see, for example, [GFA09] or [Rog+22]). As the first part of this thesis aims explicitly to solve the problem of imprecise terminology, precisely, that 'super-natural' is not a clearly defined concept, and a clear definition is missing, it would have been problematic to base such a definition on vague concepts. 'Internalizability' is, therefore, in this thesis, precisely defined as a sub-category of learnability that encompasses the acquisition of embodied procedural knowledge in the form of subconscious-level schemata that have been intentionally created by a developer or designer of the VR system. 'Congruence' is precisely defined as the degree of mapping between the long-term patterns of self and a distinct pattern of self that emerges in the use of technology, which further relates this concept to phenomenological and enactive research. Although this is disadvantageous as the application of new concepts can be more difficult in research than established and well-researched concepts, novel concepts are an important aspect of research and have the chance to advance the field, connect different domains, and produce new insights.

It can be speculated that the everyday use of VR and AR technology may become a reality in the near future, especially with smaller and less cumbersome devices. Considering the current trends of digitalization, artificial intelligence, and the proclaimed 'metaverse' [DIG13] in combination with existing hardware solutions, such as head-mounted augmented-reality displays, or proposed devices, such as retina displays [Lin+11], super-natural interaction techniques, which combine high accessibility with high potential for interacting efficiently, may someday constitute a normal aspect of our everyday life. Making the metaverse accessible to everyone and avoiding limiting technological advancements to 'magicians' [Bin00] or 'immersive natives' [Ste16a], who possess knowledge on how a modern virtuality-infused world works, can provide a challenge that needs to be addressed in parallel with technological developments. With a look at Clark's "Profiles of the Future" [Cla13], one can put it this way: *Any sufficiently established advanced technology becomes mundane.*

Super-natural interaction plays an important role in providing metaphors for VR and AR. For example, teleportation has become a conventional metaphor that is integrated into many applications. According to the proposed definition of super-natural interaction, these types of interaction can, from today's perspective, be classified as super-natural, as they have a low congruence, change the user-world relation in a beneficial way, and are highly internalizable. These techniques often begin as novel metaphors in research prototypes and projects and gradually become standards that users can easily recognize and apply, due to their utility. Super-natural interaction can thus be viewed as a paradigm in interaction design that intentionally promotes the development of techniques that are not reality-based but well-suited for VR interaction, to allow for not only more powerful technology but also more accessible and entertaining interactions.

11.3 On Immersive Telementoring

To support heart transplantation in practice, numerous technical and social challenges have to be addressed. These challenges can systematically be described using the presented ESTA framework, which is based on enactivism and activity theory. From a technical perspective, it is possible to develop immersive telementoring systems using hardware available today that are beneficial to the medical procedures. However, the design of interaction techniques remains challenging, considering the requirements of the target user group in this domain of application. Super-natural techniques, if designed well, form an approach to interaction design that can fit the requirements of hospital environments and surgeons, as well as clinicians, as the target user group.

The thesis shows that both exploration and annotation, as important tasks in immersive telementoring, can be supported using super-natural interaction techniques. In this thesis, novel interaction techniques that incorporate enactive aspects tailored for these two tasks have been investigated. The research results suggest that super-natural interaction techniques are an approach in VR design that can be advantageous in the medical domain for the following reasons: First, they can provide an effective way to approach these specific tasks and, if well-designed, perform equally well or better than naïve implementations or interactions simply transferred from the real world. Second, by reducing the required effort to learn the techniques, they can easily be acquired by surgeons, which may reduce the reluctance exhibited when VR devices are utilized in the medical domain. The prototype developed in this work addresses a gap in current research. Mobile telementoring systems are typically overlooked, as research regarding immersive telementoring often focuses on static setups with multiple sensors (see, for example, [RM+20a; Gas+21]). While both approaches have their reasonable applications and offer advantages in specific use

cases, mobile systems represent a valuable development for special use cases such as organ explantation that require options for a flexible deployment across various locations. An additional benefit of this research is that the system was tested under practical conditions, extending the research beyond phantom experiments to provide further insights. For instance, it was discovered that the camera alignment and automatic exposure settings of the HoloLens 2 are not suitable for use in operating rooms, which presumably also applies to other HMDs that integrate a forward-facing camera. This finding could positively influence the development of future devices to ensure that they meet the requirements for medical applications.

Although the developed prototype system was capable of demonstrating that many technical challenges of mobile immersive telementoring systems can be solved with today's technology, the practical application of new technologies in hospitals remains a complex process. In addition to technological advancements, cultural and social challenges must be addressed to convince decision-makers to adopt innovative solutions in healthcare settings. Regarding the ESTA model, while some contradictions have been tackled, the unaddressed contradictions must still be resolved for a complete implementation and successful transformation of health processes. Nevertheless, the development can be seen as beneficial, as demonstrators are effective in encouraging decision-makers to support innovations and introduce new processes aimed at improving the quality of care. A well-functioning system that demonstrates the desired features required in a specific use case is crucial in this context.

CHAPTER 12

CONCLUSION & FUTURE WORK

The **first part** of this thesis focuses on conceptualizing the term 'super-natural interaction.' A literature review reveals that the term lacks consistent usage in today's research, with no widely accepted definition and alternative terms that are used interchangeably (**R1**). To address this, the enactive approach was employed as a theoretical framework, which offers an interesting post-cognitivist framework that provides insights into aspects of HCI that complement other important factors analyzed within other frameworks, such as information processing theory. It focuses on how physically existing in a designed virtual world, with altered sensorimotor contingencies and unique interaction schemata, affects the user experience. Key findings from an enactive perspective, as interpreted in this thesis, emphasize that interaction design does not have to be referential to the real world or the user's real body. Instead, interaction in VR is understood as a technology-mediated engagement with the world that leads to the enacting of a technology-specific pattern of self. This emerging pattern of self can be analyzed to gain insight into the virtual existence in an antirealist virtual world and the inherent experiences (**R2**). The pattern theory of self, as well as the dimensions internalizability, congruence, and enhancement derived from this perspective on interaction, form a coherent conceptual framework capable of describing phenomenological aspects and effects on conscious experience, the ICE cube (**R3**). This complements traditional approaches in HCI research, such as quantitative comparison of performance metrics or determining effects on cognitive load.

In the **second part**, a prototype system for immersive telementoring was developed and tested under real-world conditions. Although the study does not fully explore this use case, the results suggest that mobile telementoring systems can be effectively implemented using current technology. However, with the development of a single technological experimental research prototype, the field cannot be transformed. In this regard, the presented ESTA framework facilitates the structured analysis of challenges and communication about important topics in the introduction of transformative technological intervention, both technical developments and cultural changes (**R3**). Moreover, super-natural interaction seems a promising approach for increasing the acceptance of VR systems in clinical settings. To illustrate the use of super-natural interaction in immersive telementoring, four innovative interaction techniques, ApertureZoom, PortalZoom, LensDraw, and PalmDraw have been presented and preliminarily evaluated. The conducted studies indicate that super-natural interaction can be a beneficial approach in the intended use case of heart surgery (**R5**).

This research aims to integrate diverse fields, including philosophy, design, cognitive science, HCI, and VR, into a coherent framework for HCI research. By transforming the theoretical concepts of enactivism into conceptual models, this work contributes to establishing enactivism as a complementary approach in VR and HCI research. This allows for a broader focus on additional factors that are important in contemporary HCI studies and aligns well with the concepts of third-wave [Bød15] or third-paradigm [HTS07] HCI. Enactivism offers a framework for research and a mindset for interaction design in which the user is not viewed as an information-processing system that is detached from their physical body. Instead, the user is understood as an embodied agent that actively enacts

a meaningful world within a dynamic environment and produces and constantly refines a multifaceted pattern of self in this process.

However, the topics of enactivism-informed super-natural interaction and immersive telementoring have not been fully explored, and they offer possibilities for future research. Important steps in future research could be:

- **A full evaluation of a complete system.** Both the design space and task space for annotation in immersive telementoring are by no means fully explored. In this thesis, it has only been shown that the proposed techniques can be beneficial for micro annotations. Supporting an entire operation may require other means of interaction, for example, in-air annotations, the annotation of large incisions, or the interaction with health data in VR (e.g., X-ray scans). Implementing, evaluating, and optimizing interaction techniques required for a full operation can be the next milestone after this initial experimentation.
- **Evaluating interaction with professionals.** Recruiting medical professionals for user studies can be difficult. However, for a successful introduction of super-natural interaction in VR applications intended for use in hospitals, it is essential to demonstrate that this interaction design approach is capable of enabling surgeons and other medical experts to quickly acquire the interaction schemata and apply them proficiently. With larger user studies, the preliminary findings of this thesis can be strengthened and demonstrate the application of super-natural interaction in this domain as an effective design approach.
- **Measuring internalizability, congruence, and enhancement.** The three dimensions in the ICE cube model seem promising candidates for a conceptual model that can be used to classify interaction techniques and relate different terms. However, the findings presented in this thesis can only be considered tentative. A much larger user study with more resources would be required to transform the concepts developed in this thesis into tools for VR research, such as validated questionnaires and metrics.
- **Deeper understanding of phenomenological effects.** The application in immersive telementoring was not ideal for investigating phenomenological effects. In the context of heart transplantation, both accuracy and time are the most important factors, and investigating subjective effects on the self is only of secondary importance in this context. In other contexts, such as, for example, gaming or social applications, in which performance is not the primary metric, analyzing the emerging patterns of self in depth using qualitative research methods, such as interviews, can presumably generate much deeper insights into discrete effects.
- **Presence and super-natural interaction.** The sense of presence was not explicitly addressed in this thesis. However, the literature review revealed that some researchers suspect super-natural techniques might negatively impact the feeling of presence. This, however, contrasts with the enactivism-based interpretation developed in this thesis, in which referencing reality within the agent-environment system is not considered necessary for a conscious experience of agency. Instead, alternative configurations can be viewed as equally valid foundations for consciousness. Future research may clarify these opposing conceptions.

REFERENCES

[AB91] G. D. Abowd and R. Beale. "Users, systems and interfaces: A unifying framework for interaction". In: *HCI*. Vol. 91. 1991, pp. 73–87 (Cited on pages 11, 21).

[Abt+22] P. Abtahi, S. Q. Hough, J. A. Landay, and S. Follmer. "Beyond Being Real: A Sensorimotor Control Perspective on Interactions in Virtual Reality". In: *CHI Conference on Human Factors in Computing Systems*. 2022, pp. 1–17 (Cited on pages 54, 110, 111).

[Aig+12] R. Aigner, D. Wigdor, H. Benko, M. Haller, D. Lindbauer, A. Ion, S. Zhao, and J. Koh. "Understanding mid-air hand gestures: A study of human preferences in usage of gesture types for hci". In: *Microsoft Research TechReport MSR-TR-2012-111* 2 (2012), p. 30 (Cited on page 35).

[Aks+09] N. Aksan, B. Kısac, M. Aydın, and S. Demirbuken. "Symbolic interaction theory". In: *Procedia-Social and Behavioral Sciences* 1.1 (2009), pp. 902–904 (Cited on page 129).

[All+02] R. S. Allison et al. "Simulating self-motion ii: A virtual reality tricycle". In: *Virtual Reality* 6.2 (2002), p. 86 (Cited on page 114).

[And+18] R. Ando, A. Ando, K. Kunze, and K. Minamizawa. "Bubble jumper: enhancing the traditional japanese sport sumo with physical augmentation". In: *Proceedings of the First Superhuman Sports Design Challenge: First International Symposium on Amplifying Capabilities and Competing in Mixed Realities*. 2018, pp. 1–6 (Cited on pages 60, 61).

[ACD09] A. N. Antle, G. Corness, and M. Droumeva. "Human-computer-intuition? Exploring the cognitive basis for intuition in embodied interaction". In: *International Journal of Arts and Technology* 2.3 (2009), pp. 235–254 (Cited on pages 40, 99).

[Apo+12] J. G. Apostolopoulos, P. A. Chou, B. Culbertson, T. Kalker, M. D. Trott, and S. Wee. "The road to immersive communication". In: *Proceedings of the IEEE* 100.4 (2012), pp. 974–990 (Cited on pages 61, 64).

[App23] Apple. *Introducing Apple Vision Pro*. Accessed: February 12th, 2024]. 2023. URL: `https://www.youtube.com/watch?v=TX9qSaGXFyg` (Cited on pages 3, 26).

[Arb92] M. A. Arbib. "Schema theory". In: *The encyclopedia of artificial intelligence* 2 (1992), pp. 1427–1443 (Cited on pages 36, 89).

[Arn+57] G. Arnold, F. Cahill, W. Frikell, H. L. Williams, and J. Wyman. *The Magician's Own Book, Or, The Whole Art of Conjuring*. Dick & Fitzgerald, 1857 (Cited on page 53).

[AS68] R. Atkinson and R. Shiffrin. "Human Memory: A Proposed System and its Control Processes". In: *Psychology of Learning and Motivation*. Ed. by K. W. Spence and J. T. Spence. Vol. 2. n.p.: Academic Press, 1968, pp. 89–195 (Cited on page 44).

[Aue+21] S. Auer, J. Gerken, H. Reiterer, and H.-C. Jetter. "Comparison Between Virtual Reality and Physical Flight Simulators for Cockpit Familiarization". In: *Mensch und Computer 2021*. 2021, pp. 378–392 (Cited on page 114).

[Azu97] R. T. Azuma. "A survey of augmented reality". In: *Presence: teleoperators & virtual environments* 6.4 (1997), pp. 355–385 (Cited on page 19).

[BL19] F. Bacchini and L. Lorusso. "Race, again: how face recognition technology reinforces racial discrimination". In: *Journal of information, communication and ethics in society* 17.3 (2019), pp. 321–335 (Cited on page 128).

[Bal+21] S Balakrishnan, M. S. S. Hameed, K Venkatesan, and G Aswin. "Interaction of Spatial Computing In Augmented Reality". In: *2021 7th International Conference*

on Advanced Computing and Communication Systems (ICACCS). Vol. 1. IEEE. 2021, pp. 1900–1904 (Cited on page 19).

[Ban06] A. Bandura. "Toward a psychology of human agency". In: *Perspectives on psychological science* 1.2 (2006), pp. 164–180 (Cited on pages 49, 82).

[Bañ+00] R. M. Baños, C. Botella, A. Garcia-Palacios, H. Villa, C. Perpiñá, and M. Alcaniz. "Presence and reality judgment in virtual environments: a unitary construct?" In: *CyberPsychology & Behavior* 3.3 (2000), pp. 327–335 (Cited on pages 20, 111, 123).

[BBN02] P. Barr, R. Biddle, and J. Noble. "A taxonomy of user-interface metaphors". In: *Proceedings of the SIGCHI-NZ Symposium on Computer-Human Interaction*. 2002, pp. 25–30 (Cited on pages 13, 14, 94).

[BM08] O. Bau and W. E. Mackay. "OctoPocus: a dynamic guide for learning gesture-based command sets". In: *Proceedings of the 21st annual ACM symposium on User interface software and technology*. 2008, pp. 37–46 (Cited on page 99).

[Bec+19] J. Becker, U. Meyer, T. Eichler, and S. Draheim. "A supernatural VR environment for spatial user rotation". In: *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE. 2019, pp. 850–851 (Cited on pages 60, 61, 82, 114).

[BK06] G. Bedny and W. Karwowski. *A systemic-structural theory of activity: Applications to human performance and work design*. Boca Raton, FL, USA: CRC press, 2006 (Cited on page 50).

[BH95] Y. Benjamini and Y. Hochberg. "Controlling the false discovery rate: a practical and powerful approach to multiple testing". In: *Journal of the Royal statistical society: series B (Methodological)* 57.1 (1995), pp. 289–300 (Cited on page 149).

[Ben93] D. Benyon. "Adaptive systems: a solution to usability problems". In: *User modeling and User-adapted Interaction* 3.1 (1993), pp. 65–87 (Cited on page 104).

[Ber+08] D. Bernstein, L. Penner, A Clarke-Stewart, and E. Roy. *Psychology (8. izd.)* 2008 (Cited on pages 14, 15, 36, 40, 42, 44, 104).

[BB03] O. W. Bertelsen and S. Bødker. "Activity theory". In: San Fransisco: Morgan Kaufmann, 2003, pp. 291–324 (Cited on pages 49, 50).

[BP93] P. A. Bibby and S. J. Payne. "Internalizing and the use specificity of device knowledge". In: *Human-computer interaction* 8.1 (1993), pp. 25–56 (Cited on page 99).

[Bie+93] E. A. Bier, M. C. Stone, K. Pier, W. Buxton, and T. D. DeRose. "Toolglass and magic lenses: the see-through interface". In: *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*. 1993, pp. 73–80 (Cited on page 155).

[Bin00] K. Binsted. "Sufficiently advanced technology: using magic to control the world". In: *CHI'00 Extended Abstracts on Human Factors in Computing Systems*. 2000, pp. 205–206 (Cited on pages 71, 183).

[Bir+22] M. Birlo, P. E. Edwards, M. Clarkson, and D. Stoyanov. "Utility of optical see-through head mounted displays in augmented reality-assisted surgery: A systematic review". In: *Medical Image Analysis* 77 (2022) (Cited on pages 131, 136).

[Bla11] J. Blake. *Natural user interfaces in. NET: WPF 4, Surface 2, and Kinect*. Manning, 2011 (Cited on page 17).

[Bla+90] C. Blanchard, S. Burgess, Y. Harvill, J. Lanier, A. Lasko, M. Oberman, and M. Teitel. "Reality built for two: a virtual reality tool". In: *Proceedings of the 1990 symposium on Interactive 3D graphics*. 1990, pp. 35–36 (Cited on page 26).

[Blo+86] E. Bloch, N. Plaice, S. Plaice, and P. Knight. *The principle of hope*. Vol. 2. mit Press Cambridge, MA, 1986 (Cited on page 71).

[BX12] H. Blumenthal and Y. Xu. "The ghost club storyscape: designing for transmedia storytelling". In: *IEEE Transactions on Consumer Electronics* 58.2 (2012), pp. 190–196 (Cited on page 61).

[Bø06] S. Bødker. "When Second Wave HCI Meets Third Wave Challenges". In: *Proceedings of the 4th Nordic Conference on Human-Computer Interaction: Changing Roles*. NordiCHI '06. Oslo, Norway: Association for Computing Machinery, 2006, pp. 1–8. ISBN: 1595933255 (Cited on page 129).

[Bød15] S. Bødker. "Third-wave HCI, 10 years later—participation and sharing". In: *interactions* 22.5 (2015), pp. 24–31 (Cited on pages 181, 185).

[BH92] R. A. Bolt and E. Herranz. "Two-handed gesture in multi-modal natural dialog". In: *Proceedings of the 5th annual ACM symposium on User interface software and technology*. 1992, pp. 7–14 (Cited on page 34).

[BB19] A. Börütecene and O. Buruk. "Otherworld: Ouija Board as a Resource for Design". In: *Proceedings of the Halfway to the Future Symposium 2019*. 2019, pp. 1–4 (Cited on page 61).

[BMR12] D. Bowman, R. McMahan, and E. Ragan. "Questioning Naturalism in 3D User Interfaces". In: *Communications of The ACM - CACM* 55 (Sept. 2012) (Cited on pages 22, 54, 56, 68).

[BH97] D. A. Bowman and L. F. Hodges. "An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments". In: *Proceedings of the 1997 symposium on Interactive 3D graphics*. 1997, 35–ff (Cited on pages 30, 31, 93).

[BH99] D. A. Bowman and L. F. Hodges. "Formalizing the design, evaluation, and application of interaction techniques for immersive virtual environments". In: *Journal of Visual Languages & Computing* 10.1 (1999), pp. 37–53 (Cited on pages 13, 28, 31, 90).

[BKH97] D. A. Bowman, D. Koller, and L. F. Hodges. "Travel in immersive virtual environments: An evaluation of viewpoint motion control techniques". In: *Proceedings of IEEE 1997 Annual International Symposium on Virtual Reality*. IEEE. 1997, pp. 45–52 (Cited on pages 28–30).

[Box76] G. E. Box. "Science and statistics". In: *Journal of the American Statistical Association* 71.356 (1976), pp. 791–799 (Cited on page 116).

[Boz+16] E. Bozgeyikli, A. Raij, S. Katkoori, and R. Dubey. "Point & teleport locomotion technique for virtual reality". In: *Proceedings of the 2016 annual symposium on computer-human interaction in play*. 2016, pp. 205–216 (Cited on pages 30, 70, 76, 82, 84, 85, 94, 97, 106, 114, 115, 117).

[Bre56] D. Brewster. *The Stereoscope; Its History, Theory, and Construction: with Its Application to the Fine and Useful Arts and to Education*. John Murray, 1856 (Cited on page 23).

[Bre05] P. Brey. "The epistemology and ontology of human-computer interaction". In: *Minds and Machines* 15.3-4 (2005), pp. 383–398 (Cited on pages 74, 88).

[Bro+96] J. Brooke et al. "SUS - A quick and dirty usability scale". In: *Usability evaluation in industry* 189.194 (1996), pp. 4–7 (Cited on page 230).

[Bro99] F. P. Brooks. "What's real about virtual reality?" In: *IEEE Computer graphics and applications* 19.6 (1999), pp. 16–27 (Cited on page 18).

[BRIM14] P. C. Brown, H. L. Roediger III, and M. A. McDaniel. *Make it stick: The science of successful learning*. Harvard University Press, 2014 (Cited on page 16).

[Bry05] S. Bryson. "Direct Manipulation in Virtual Reality". In: *The Visualization Handbook, Charles D. Hansen & Chris R. Johnson, 2005, Elsevier Inc* (2005), p. 413 (Cited on page 35).

[Buc+04] V. Buchmann, S. Violich, M. Billinghurst, and A. Cockburn. "FingARtips: gesture based direct manipulation in Augmented Reality". In: *Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Australasia and South East Asia*. 2004, pp. 212–221 (Cited on page 35).

[Buc+18] T. Buckers, B. Gong, E. Eisemann, and S. Lukosch. "VRabl: stimulating physical activities through a multiplayer augmented reality sports game". In: *Proceedings of the First Superhuman Sports Design Challenge: First International Symposium on Amplifying Capabilities and Competing in Mixed Realities*. 2018, pp. 1–5 (Cited on pages 60, 61).

[BDPB13] T. Buhrmann, E. A. Di Paolo, and X. Barandiaran. "A dynamical systems account of sensorimotor contingencies". In: *Frontiers in psychology* 4 (2013), p. 285 (Cited on pages 91, 92).

[Bul13] J. Bullington. *The expression of the psychosomatic body from a phenomenological perspective*. Springer, 2013 (Cited on page 72).

[BB20] E. Burches and M. Burches. "Efficacy, effectiveness and efficiency in the health care: The need for an agreement to clarify its meaning". In: *Int Arch Public Health Community Med* 4.1 (2020), pp. 1–3 (Cited on page 128).

[BC03] G. C. Burdea and P. Coiffet. *Virtual reality technology*. John Wiley & Sons, 2003 (Cited on page 24).

[Bur+19] S. M. van der Burgt, R. A. Kusurkar, J. A. Wilschut, T. A. Tsoi, L. Sharon, G. Croiset, and S. M. Peerdeman. "Medical specialists' basic psychological needs, and motivation for work and lifelong learning: a two-step factor score path analysis". In: *BMC medical education* 19.1 (2019), pp. 1–11 (Cited on pages 131, 135).

[Büs+19] W. Büschel, A. Mitschick, T. Meyer, and R. Dachselt. "Investigating Smartphone-based Pan and Zoom in 3D Data Spaces in Augmented Reality". In: (2019) (Cited on page 155).

[Bux83] W. Buxton. "Lexical and pragmatic considerations of input structures". In: *ACM SIGGRAPH Computer Graphics* 17.1 (1983), pp. 31–37 (Cited on page 11).

[Byr+22] D. Byrne et al. "Spooky Technology: The ethereal and otherworldly as a resource for design". In: *Designing Interactive Systems Conference*. 2022, pp. 759–775 (Cited on page 61).

[BMM16] R. Byrne, J. Marshall, and F. Mueller. "Balance ninja: towards the design of digital vertigo games via galvanic vestibular stimulation". In: *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*. 2016, pp. 159–170 (Cited on page 61).

[Cad04] C. Cadoz. *Enactive Interfaces? by Claude Cadoz*. 2004 (Cited on page 48).

[CG17] E. Camina and F. Güell. "The neuroanatomical, neurophysiological and psychological basis of memory: Current models and their origins". In: *Frontiers in pharmacology* 8 (2017), p. 438 (Cited on pages 41, 43, 44, 99).

[Can22] D. Cantone. "The simulated body: A preliminary investigation into the relationship between neuroscientific studies, phenomenology and virtual reality". In: *Foundations of Science* (2022), pp. 1–10 (Cited on pages 48, 88).

[CW12] M. Cappuccio and M. Wheeler. "Ground-level intelligence: Action-oriented representation and the dynamics of the background". In: *Knowing without Thinking: Mind, Action, Cognition and the Phenomenon of the Background*. Springer, 2012, pp. 13–36 (Cited on page 89).

[CM88] S. K. Card and T. P. Moran. "User technology: From pointing to pondering". In: *A history of personal workstations*. 1988, pp. 489–526 (Cited on pages 14, 15).

[CNM83] S. K. Card, A. Newell, and T. P. Moran. *The Psychology of Human-Computer Interaction*. 1983 (Cited on pages 40–42, 88–90, 93, 105).

[CMN86] S. Card, T. MORAN, and A. Newell. "The model human processor- An engineering model of human performance". In: *Handbook of perception and human performance*. 2.45–1 (1986) (Cited on pages 103, 105).

[Car03] J. M. Carroll. *HCI models, theories, and frameworks: Toward a multidisciplinary science*. Elsevier, 2003 (Cited on pages 39, 40).

[CMK88] J. M. Carroll, R. L. Mack, and W. A. Kellogg. "Interface metaphors and user interface design". In: *Handbook of human-computer interaction*. Elsevier, 1988, pp. 67–85 (Cited on pages 13, 14).

[CO88] J. M. Carroll and J. R. Olson. "Mental models in human-computer interaction". In: *Handbook of human-computer interaction* (1988), pp. 45–65 (Cited on pages 14, 104).

[Cas+21] F. A. Casari et al. "Augmented reality in orthopedic surgery is emerging from proof of concept towards clinical studies: a literature review explaining the technology and current state of the art". In: *Current Reviews in Musculoskeletal Medicine* 14.2 (2021), pp. 192–203 (Cited on page 131).

[CHC15] P. Cash, B. Hicks, and S. Culley. "Activity Theory as a means for multi-scale analysis of the engineering design process: A protocol study of design in practice". In: *Design Studies* 38 (2015), pp. 1–32 (Cited on page 50).

[CWLJ12] J. Cashion, C. Wingrave, and J. J. LaViola Jr. "Dense and dynamic 3d selection for game-based virtual environments". In: *IEEE transactions on visualization and computer graphics* 18.4 (2012), pp. 634–642 (Cited on page 155).

[CRV12] G. Casiez, N. Roussel, and D. Vogel. "1€ filter: a simple speed-based low-pass filter for noisy input in interactive systems". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2012, pp. 2527–2530 (Cited on page 158).

[CF08] A. Chalmers and A. Ferko. "Levels of realism: From virtual reality to real virtuality". In: *Proceedings of the 24th Spring Conference on Computer Graphics*. 2008, pp. 19–25 (Cited on page 21).

[CC17] H. Chang and M. F. Cohen. "Panning and zooming high-resolution panoramas in virtual reality devices". In: *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*. 2017, pp. 279–288 (Cited on page 155).

[Che11] A. Chemero. *Radical Embodied Cognitive Science*. MIT Press, 2011 (Cited on page 81).

[Chi+81] M. T. H. Chi et al. *Expertise in Problem Solving*. ERIC, 1981 (Cited on pages 36, 39, 89).

[CS01] L. Chittaro and I. Scagnetto. "Is semitransparency useful for navigating virtual environments?" In: *Proceedings of the ACM symposium on Virtual reality software and technology*. 2001, pp. 159–166 (Cited on pages 60, 61).

[Cho+22] Y. Choi, H. Jeon, S. Lee, I. Han, Y. Luo, S. Kim, W. Matusik, and K. Kim. "Seamless-walk: Novel Natural Virtual Reality Locomotion Method with a High-Resolution Tactile Sensor". In: *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE. 2022, pp. 696–697 (Cited on page 115).

[Cla04] A. Clark. *The Twisted Matrix: Dream, Simulation or Hybrid?* Oxford University Press, 2004 (Cited on page 80).

[Cla08] A. Clark. *Supersizing the mind: Embodiment, action, and cognitive extension*. Oxfort University Press, 2008 (Cited on page 100).

[Cla13] A. C. Clarke. *Profiles of the Future*. Hachette UK, 2013 (Cited on page 183).

[Cle14] T. P. Clement. "Authorship matrix: a rational approach to quantify individual contributions and responsibilities in multi-author scientific articles". In: *Science and engineering ethics* 20.2 (2014), pp. 345–361 (Cited on page 221).

[CiN16] T. Clemmensen, V. in, and B. Nardi. "Making HCI theory work: an analysis of the use of activity theory in HCI research". In: *Behaviour & Information Technology* 35.8 (2016), pp. 608–627 (Cited on pages 49, 50).

[CS14] J. Cogburn and M. Silcox. "Against brain-in-a-vatism: On the value of virtual reality". In: *Philosophy & Technology* 27 (2014), pp. 561–579 (Cited on pages 48, 88).

[Coh+85] N. J. Cohen, H. Eichenbaum, B. S. Deacedo, and S. Corkin. "Different memory systems underlying acquisition of procedural and declarative knowledge." In: *Annals of the New York Academy of Sciences* (1985) (Cited on page 103).

[Col20] I. G. Colditz. "A consideration of physiological regulation from the perspective of Bayesian enactivism". In: *Physiology & behavior* 214 (2020), p. 112758 (Cited on page 44).

[Col14] G. Colombetti. "The feeling body". In: *Affective Science Meets the Enactive Mind* (2014) (Cited on page 46).

[Com14] Computer History Museum. *Odysseys in Technology: Research and Fun, lecture by Ivan Sutherland*. Date of Recording: October 19th 2005. [Accessed: January 17th, 2023]. 2014. URL: `https://www.youtube.com/watch?v=FIMaf4RemOU` (Cited on page 23).

[CS21] R. Cools and A. Simeone. "Mobile Displays for Cross-Reality Interactions between Virtual and Physical Realities". In: *Proceedings of the 20th International Conference on Mobile and Ubiquitous Multimedia*. 2021, pp. 217–219 (Cited on pages 60, 61).

[Cos+19] W. Costa, L. Ananias, I. Barbosa, B. Barbosa, A. De'Carli, R. R. Barioni, L. Figueiredo, V. Teichrieb, and D. Filgueira. "Songverse: a music-loop authoring tool based on Virtual Reality". In: *2019 21st Symposium on Virtual and Augmented Reality (SVR)*. IEEE. 2019, pp. 216–222 (Cited on page 61).

[CN+92] C. Cruz-Neira, D. J. Sandin, T. A. DeFanti, R. V. Kenyon, and J. C. Hart. "The CAVE: audio visual experience automatic virtual environment". In: *Communications of the ACM* 35.6 (1992), pp. 64–73 (Cited on pages 2, 24).

[Csi14] M. Csikszentmihalyi. "Play and intrinsic rewards". In: *Flow and the foundations of positive psychology: The collected works of Mihaly Csikszentmihalyi* (2014), pp. 135–153 (Cited on page 91).

[Cue17] E. Cuervo. "Beyond reality: Head-mounted displays for mobile systems researchers". In: *GetMobile: Mobile Computing and Communications* 21.2 (2017), pp. 9–15 (Cited on page 26).

[DAA98] R. P. Darken, T. Allard, and L. B. Achille. "Spatial orientation and wayfinding in large-scale virtual spaces: An introduction". In: *Presence* 7.2 (1998), pp. 101–107 (Cited on page 29).

[DF73] T. A. De Fanti. "The Graphics Symbiosis System–An Interactive Mini-Computer Animation Graphics Language Designed for Habitability and Extensibility". PhD thesis. The Ohio State University, 1973 (Cited on pages 17, 68, 69).

[DJ16] P. De Jesus. "Autopoietic enactivism, phenomenology and the deep continuity between life and mind". In: *Phenomenology and the Cognitive Sciences* 15 (2016), pp. 265–289 (Cited on page 129).

[DO17] J. Degenaar and J. K. O'regan. "Sensorimotor theory and enactivism". In: *Topoi* 36 (2017), pp. 393–407 (Cited on page 48).

[Den88] D. C. Dennett. "Quining qualia". In: *Consciousness in contemporary science* (1988), pp. 42–77 (Cited on page 20).

[Den+21] C. Dennler, D. E. Bauer, A.-G. Scheibler, J. Spirig, T. Götschi, P. Fürnstahl, and M. Farshad. "Augmented reality in the operating room: A clinical feasibility study". In: *BMC musculoskeletal disorders* 22.1 (2021), pp. 1–9 (Cited on page 131).

[Der96] S. J. Derry. "Cognitive schema theory in the constructivist debate". In: *Educational psychologist* 31.3-4 (1996), pp. 163–174 (Cited on pages 37, 105).

[Des21] M. Destéfano. "Cognitivism and the intellectualist vision of the mind". In: *Phenomenology and Mind* 21 (2021), pp. 142–153 (Cited on page 48).

[Dew+22a] B. Dewitz, R. Bibo, S. Kalkhoff, S. Moazemi, A. Liebrecht, C. Geiger, F. Steinicke, H. Aubin, and F. Schmid. *Towards 5G Telementoring in VR-Assisted Heart Trans-*

plantation Using HoloLens 2. GI VR / AR Workshop 2022. 2022 (Cited on pages 139, 221).

[DGS21] B. Dewitz, C. Geiger, and F. Steinicke. "Virtual Visus-Vision Acuity and Text Legibility in Virtual Environments". In: *GI VR/AR Workshop*. Gesellschaft für Informatik eV. 2021 (Cited on page 221).

[DGS22] B. Dewitz, C. Geiger, and F. Steinicke. "Acting Beyond Reality –- The Role of Schemata in Mixed-Reality Super-Natural Interaction". In: *GI VR/AR Workshop*. Gesellschaft für Informatik eV. 2022 (Cited on page 221).

[DGS23] B. Dewitz, C. Geiger, and F. Steinicke. "Enactinv Interaction in Virtual Reality". In: *GI VR/AR Workshop*. Gesellschaft für Informatik eV. 2023 (Cited on pages 45, 80, 221).

[Dew+21] B. Dewitz, C. Geiger, F. Steinicke, and C. Huhn. "Virtuality between my Fingers– Investigation of Zoom Mechanisms for Visual Exploration of Virtual Environments". In: *GI VR/AR Workshop*. Gesellschaft für Informatik eV. 2021 (Cited on page 221).

[DGS20] B. Dewitz, C. Geiger, and F. Steinicke. "Hand-Based Interaction on a Millimeter Scale in Virtual and Augmented Reality". In: *GI VR/AR Workshop*. Gesellschaft für Informatik eV. 2020 (Cited on pages 27, 221).

[Dew+23a] B. Dewitz, S. Karaosmanoglu, R. W. Lindeman, and F. Steinicke. "Magic, Superpowers, or Empowerment? A Conceptual Framework for Magic Interaction Techniques". In: *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 2023, pp. 807–808 (Cited on page 221).

[Dew+18] B. Dewitz, P. Ladwig, F. Steinicke, and C. Geiger. "Classification of Beyond-Reality Interaction Techniques in Spatial Human-Computer Interaction". In: *Proceedings of the Symposium on Spatial User Interaction*. 2018, pp. 185–185 (Cited on pages 4, 221).

[Dew+23b] B. Dewitz, S. Moazemi, S. Kalkhoff, S. Kessler, C. Geiger, F. Steinicke, H. Aubin, and F. Schmid. "Enacted Selves in Technological Activities – Framework and Case Study in Immersive Telementoring". In: *Mensch Und Computer 2023*. MuC '23. Rapperswil, Switzerland: Association for Computing Machinery, 2023, pp. 289–299 (Cited on pages 50, 126, 134, 221).

[DSG19] B. Dewitz, F. Steinicke, and C. Geiger. "Functional workspace for one-handed tap and swipe microgestures". In: *Mensch und Computer 2019-Workshopband* (2019) (Cited on pages 33, 221).

[Dew+22b] B. Dewitz et al. "Real-time 3D scans of cardiac surgery using a single optical-see-through head-mounted display in a mobile setup". In: *Frontiers in Virtual Reality* 3 (2022) (Cited on pages 147, 151, 221).

[DPBB17] E. Di Paolo, T. Buhrmann, and X. Barandiaran. *Sensorimotor life: An enactive proposal*. Oxford University Press, 2017 (Cited on pages 36, 38, 44–48, 88–92, 95, 96, 98, 179).

[DPRDJ10] E. Di Paolo, M. Rohde, and H. De Jaegher. "Horizons for the enactive mind: Values, social interaction, and play". In: *Enaction: Towards a new paradigm for cognitive science*. 2010 (Cited on page 46).

[DP05] E. A. Di Paolo. "Autopoiesis, adaptivity, teleology, agency". In: *Phenomenology and the cognitive sciences* 4.4 (2005), pp. 429–452 (Cited on page 45).

[DPCDJ18] E. A. Di Paolo, E. C. Cuffari, and H. De Jaegher. *Linguistic bodies: The continuity between life and language*. MIT press, 2018 (Cited on pages 45, 47).

[DE10] B. C. Dickerson and H. Eichenbaum. "The episodic memory system: neurocircuitry and disorders". In: *Neuropsychopharmacology* 35.1 (2010), pp. 86–104 (Cited on page 43).

[Dil71] M. C. Dillon. "Gestalt Theory and Merleau-Ponty's Concept of Intentionality". In: *Man and World* 4.4 (1971), pp. 436–459 (Cited on page 72).

[DIG13] J. D. N. Dionisio, W. G. B. III, and R. Gilbert. "3D virtual worlds and the metaverse: Current status and future possibilities". In: *ACM Computing Surveys (CSUR)* 45.3 (2013), pp. 1–38 (Cited on pages 114, 183).

[Dix+04] A. Dix, J. Finlay, G. D. Abowd, and R. Beale. *Human-computer interaction*. Pearson Education, 2004 (Cited on pages 10, 16, 103, 106).

[Doe+22] R. Doerner, W. Broll, P. Grimm, and B. Jung. *Virtual and augmented reality (VR/AR): Foundations and methods of extended realities (XR)*. Springer Nature, 2022 (Cited on pages 19, 23–26).

[Dou04] P. Dourish. *Where the action is: the foundations of embodied interaction*. MIT press, 2004 (Cited on pages 13, 80, 81, 88, 95).

[Dre96] H. L. Dreyfus. "The current relevance of Merleau-Ponty's phenomenology of embodiment". In: *The electronic journal of analytic philosophy* 4.4 (1996), pp. 1–16 (Cited on pages 81, 89).

[Dre14] H. L. Dreyfus. *Skillful coping: Essays on the phenomenology of everyday perception and action*. Oxford University Press, 2014 (Cited on page 47).

[DEN16] F. Dufour and C. Ehrwein Nihan. "Do robots need to be stereotyped? Technical characteristics as a moderator of gender stereotyping". In: *Social sciences* 5.3 (2016), p. 27 (Cited on page 128).

[EWK18] C. Eghtebas, S. Weber, and G. Klinker. "Investigation into Natural Gestures Using EMG for "SuperNatural" Interaction in VR". In: *Adjunct Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*. 2018, pp. 102–104 (Cited on pages 56, 60, 61, 71).

[EM94] T. Eiter and H. Mannila. *Computing discrete Fréchet distance*. 1994 (Cited on page 172).

[Ell+14] R. D. Ellis et al. "Enactivism and the New Teleology: Reconciling the Warring Camps". In: *AVANT. Pismo Awangardy Filozoficzno-Naukowej* 2 (2014), pp. 173–198 (Cited on page 44).

[Ell91] S. R. Ellis. "Varieties of virtualization". In: *Human Machine Interfaces for Teleoperators and Virtual Environments* 10071 (1991), p. 78 (Cited on page 54).

[Eng23] D. C. Engelbart. "Augmenting human intellect: A conceptual framework". In: *Augmented Education in the Global Age*. Routledge, 2023, pp. 13–29 (Cited on page 10).

[EE68] D. C. Engelbart and W. K. English. "A research center for augmenting human intellect". In: *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*. 1968, pp. 395–410 (Cited on page 11).

[Eng01] Y. Engeström. "Expansive learning at work: Toward an activity theoretical reconceptualization". In: *Journal of education and work* 14.1 (2001), pp. 133–156 (Cited on pages 49, 50, 125).

[Eri+19] A. Erickson, K. Kim, R. Schubert, G. Bruder, and G. Welch. "Is it cold in here or is it just me? analysis of augmented reality temperature visualization for computer-mediated thermoception". In: *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE. 2019, pp. 202–211 (Cited on page 112).

[EHK18] K. A. Ericsson, R. R. Hoffman, and A. Kozbelt. *The Cambridge handbook of expertise and expert performance*. Cambridge University Press, 2018 (Cited on page 43).

[Err+19] S. Erridge, D. K. Yeung, H. R. Patel, and S. Purkayastha. "Telementoring of Surgeons: A Systematic Review". In: *Surgical Innovation* 26.1 (2019), pp. 95–111 (Cited on page 130).

[Eva03] J. S. B. Evans. "In two minds: dual-process accounts of reasoning". In: *Trends in cognitive sciences* 7.10 (2003), pp. 454–459 (Cited on pages 41, 179).

[Far94] S. M. Faris. "Novel 3D stereoscopic imaging technology". In: *Stereoscopic Displays and Virtual Reality Systems*. Vol. 2177. SPIE. 1994, pp. 180–195 (Cited on page 23).

[Fei+15] T. Feix, J. Romero, H.-B. Schmiedmayer, A. M. Dollar, and D. Kragic. "The grasp taxonomy of human grasp types". In: *IEEE Transactions on human-machine systems* 46.1 (2015), pp. 66–77 (Cited on pages 33, 97).

[Fel15] V. J. Feltham. *Palmer Luckey Explains Oculus Rift's Constellation Tracking and Fabric*. https://web.archive.org/web/20190314101939/https://www.vr-focus. com/2015/06/palmer-luckey-explains-oculus-rifts-constellation-tracking-and-fabric [Accessed: June 30th, 2023]. 2015 (Cited on page 26).

[Fis80] K. W. Fischer. "A theory of cognitive development: The control and construction of hierarchies of skills." In: *Psychological review* 87.6 (1980), p. 477 (Cited on pages 37, 105).

[Fis+22] R. Fischer, A. Mühlenbrock, F. Kulapichitr, V. N. Uslar, D. Weyhe, and G. Zachmann. "Evaluation of Point Cloud Streaming and Rendering for VR-Based Telepresence in the OR". In: *Virtual Reality and Mixed Reality*. Ed. by G. Zachmann, M. Alcañiz Raya, P. Bourdot, M. Marchal, J. Stefanucci, and X. Yang. Cham: Springer International Publishing, 2022, pp. 89–110 (Cited on page 131).

[Fis+17] J. A. Fisher, A. Garg, K. P. Singh, and W. Wang. "Designing intentional impossible spaces in virtual reality narratives: A case study". In: *2017 IEEE Virtual Reality (VR)*. IEEE. 2017, pp. 379–380 (Cited on pages 60, 61, 64).

[FP67] P. M. Fitts and M. I. Posner. *Human performance.* Brooks/Cole, 1967 (Cited on page 44).

[FWC84] J. D. Foley, V. L. Wallace, and P. Chan. "The human factors of computer graphics interaction techniques". In: *IEEE computer Graphics and Applications* 4.11 (1984), pp. 13–48 (Cited on pages 12, 181).

[For+68] J. W. Forrester et al. *Principles of systems*. Wright-Allen Press Cambridge, MA, 1968 (Cited on page 15).

[FHZ96] A. Forsberg, K. Herndon, and R. Zeleznik. "Aperture based selection for immersive virtual environments". In: *Proceedings of the 9th annual ACM symposium on User interface software and technology*. 1996, pp. 95–96 (Cited on pages 31, 155).

[For+00] A. S. Forsberg, D. H. Laidlaw, A. Van Dam, R. M. Kirby, G. Kafniadakis, and J. L. Elion. "Immersive virtual reality for visualizing flow through an artery". In: *Proceedings Visualization 2000. VIS 2000 (Cat. No. 00CH37145)*. IEEE. 2000, pp. 457–460 (Cited on page 165).

[FAW19] T. Franke, C. Attig, and D. Wessel. "A personal resource for technology interaction: development and validation of the affinity for technology interaction (ATI) scale". In: *International Journal of Human–Computer Interaction* 35.6 (2019), pp. 456–467 (Cited on page 126).

[Fra19] C. Frauenberger. "Entanglement HCI the next wave?" In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 27.1 (2019), pp. 1–27 (Cited on pages 72, 129).

[Fre73] D. Freedman. "Computer Magic". In: *Proceedings of the eleventh annual SIGCPS computer personnel research conference*. 1973, pp. 1–9 (Cited on page 53).

[Fre+20] J. P. Freiwald, O. Ariza, O. Janeh, and F. Steinicke. "Walking by Cycling: A Novel In-Place Locomotion User Interface for Seated Virtual Reality Experiences." In: *CHI*. 2020, pp. 1–12 (Cited on pages 98, 115).

[Fre10] S. Freud. *Civilization and Its Discontents*. W.W. Norton, 2010 (Cited on page 71).

[FDP11] T. Froese and E. A. Di Paolo. "The enactive approach: Theoretical sketches from cell to society". In: *Pragmatics & Cognition* 19.1 (2011), pp. 1–36 (Cited on pages 47, 90, 91, 102).

[Fu+18] L. P. Fu, J. Landay, M. Nebeling, Y. Xu, and C. Zhao. "Redefining natural user interface". In: *Extended abstracts of the 2018 CHI conference on human factors in computing systems*. 2018, pp. 1–3 (Cited on page 108).

[Fur+19] T. Furumoto, M. Ito, M. Fujiwara, Y. Makino, H. Shinoda, and T. Kamigaki. "Three-dimensional interaction technique using an acoustically manipulated balloon". In: *SIGGRAPH Asia 2019 Emerging Technologies*. 2019, pp. 51–52 (Cited on page 61).

[Gad+14] V. R. Gaddam, R. Langseth, S. Ljødal, P. Gurdjos, V. Charvillat, C. Griwodz, and P. Halvorsen. "Interactive zoom and panning from live panoramic video". In: *Proceedings of Network and Operating System Support on Digital Audio and Video Workshop*. 2014, pp. 19–24 (Cited on page 155).

[GDA19] T. Galais, A. Delmas, and R. Alonso. "Natural Interaction in Virtual Reality: Impact on the Cognitive Load". In: *Adjunct Proceedings of the 31st Conference on l'Interaction Homme-Machine*. IHM '19 Adjunct. Grenoble, France: Association for Computing Machinery, 2019 (Cited on page 171).

[Gal+20] R. Galati, M. Simone, G. Barile, R. De Luca, C. Cartanese, and G Grassi. "Experimental setup employed in the operating room based on virtual and mixed reality: analysis of pros and cons in open abdomen surgery". In: *Journal of healthcare engineering* 2020 (2020) (Cited on pages 130, 132).

[Gal86] S. Gallagher. "Body image and body schema: A conceptual clarification". In: *The Journal of mind and behavior* (1986), pp. 541–554 (Cited on pages 36, 80).

[Gal08] S. Gallagher. "Are minimal representations still representations?" In: *International Journal of Philosophical Studies* 16.3 (2008), pp. 351–369 (Cited on page 89).

[Gal13] S. Gallagher. "A pattern theory of self". In: *Frontiers in human neuroscience* 7 (2013), p. 443 (Cited on pages 6, 74, 75, 128, 179).

[Gal17] S. Gallagher. *Enactivist interventions: Rethinking the mind*. Oxford University Press, 2017 (Cited on pages 47, 89).

[Gal22] S. Gallagher. "What Is Phenomenology?" In: *Phenomenology*. Cham: Springer International Publishing, 2022, pp. 1–10 (Cited on page 72).

[Gao+22] K. Gao, Y. Gao, H. He, D. Lu, L. Xu, and J. Li. "Nerf: Neural radiance field in 3d vision, a comprehensive review". In: *arXiv preprint arXiv:2210.00379* (2022) (Cited on page 152).

[GA04] F. Garbarini and M. Adenzato. "At the root of embodied cognition: Cognitive science meets neurophysiology". In: *Brain and cognition* 56.1 (2004), pp. 100–106 (Cited on page 89).

[Gar+18] B. Garrett, T. Taverner, D. Gromala, G. Tao, E. Cordingley, C. Sun, et al. "Virtual reality clinical research: promises and challenges". In: *JMIR serious games* 6.4 (2018), e10839 (Cited on page 125).

[Gas+21] D. Gasques et al. "ARTEMIS: A Collaborative Mixed-Reality System for Immersive Surgical Telementoring". In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. CHI '21. Yokohama, Japan: Association for Computing Machinery, 2021, pp. 1–14 (Cited on pages 130, 140, 183).

[Gau+14] S. Gauglitz, B. Nuernberger, M. Turk, and T. Höllerer. "World-stabilized annotations and virtual scene navigation for remote collaboration". In: *Proceedings of the 27th annual ACM symposium on User interface software and technology*. 2014, pp. 449–459 (Cited on pages 60, 61).

[Geh80] A. Gehlen. *Man in the Age of Technology*. New York: Columbia University Press, 1980 (Cited on page 71).

[GR88] A. Gehlen and K.-S. Rehberg. *Man, his nature and place in the world*. Vol. 3. Columbia University Press, 1988 (Cited on pages 71, 112).

[Gib14] J. J. Gibson. *The ecological approach to visual perception: classic edition*. Psychology press, 2014 (Cited on pages 15, 81).

[Gie+18] A. Giesbrecht, S. von Styp-Rekowski, B. Dewitz, and C. Geiger. "Examining effects of altered gravity direction in Room-Scale VR". In: *GI VR/AR Workshop*. Gesellschaft für Informatik eV. 2018 (Cited on pages 70, 82, 114, 221).

[Gil90] K. J. Gilhooly. "Cognitive psychology and medical diagnosis". In: *Applied cognitive psychology* 4.4 (1990), pp. 261–272 (Cited on page 36).

[GP90] D. A. Gioia and E. Pitre. "Multiparadigm perspectives on theory building". In: *Academy of management review* 15.4 (1990), pp. 584–602 (Cited on page 181).

[Gis+19] M. van Gisbergen, M. Kovacs, F. Campos, M. van der Heeft, and V. Vugts. "What we don't know. the effect of realism in virtual reality on experience and behaviour". In: *Augmented Reality and Virtual Reality: The Power of AR and VR for Business* (2019), pp. 45–57 (Cited on page 21).

[Gla19] M. E. Gladden. "Novel forms of "magical" human-computer interaction within the cyber-physical smart workplace: Implications for usability and user experience". In: *International Journal of Research Studies in Management* 8.1 (2019), pp. 25–48 (Cited on pages 55, 66).

[Glo+14] D. Glomberg, D. Kirchhof, O. Köse, F. Schöndorff, M. Tiator, R. Wiche, and C. Geiger. "ZeroGravity-eine virtuelle Nutzererfahrung in Luft und Wasser". In: *Mensch & Computer 2014-Workshopband* (2014) (Cited on page 114).

[GC21] E. B. Goldstein and L. Cacciamani. *Sensation and perception*. Cengage Learning, 2021 (Cited on pages 40, 44).

[GFL17] M. Gonzalez-Franco and J. Lanier. "Model of illusions and virtual reality". In: *Frontiers in psychology* 8 (2017), p. 1125 (Cited on page 20).

[Gra+18] A. Granqvist, T. Takala, J. Takatalo, and P. Hämäläinen. "Exaggeration of avatar flexibility in virtual reality". In: *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*. 2018, pp. 201–209 (Cited on page 56).

[GFA09] T. Grossman, G. Fitzmaurice, and R. Attar. "A survey of software learnability: metrics, methodologies and guidelines". In: *Proceedings of the sigchi conference on human factors in computing systems*. 2009, pp. 649–658 (Cited on pages 16, 182).

[GBB20] U. Gruenefeld, Y. Brück, and S. Boll. "Behind the scenes: Comparing x-ray visualization techniques in head-mounted optical see-through augmented reality". In: *Proceedings of the 19th International Conference on Mobile and Ubiquitous Multimedia*. 2020, pp. 179–185 (Cited on pages 60, 61).

[Gsa+19] C. Gsaxner, A. Pepe, J. Wallner, D. Schmalstieg, and J. Egger. "Markerless image-to-face registration for untethered augmented reality in head and neck surgery". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2019, pp. 236–244 (Cited on page 132).

[Gug+16] 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*. 2016, pp. 49–60 (Cited on pages 61, 63).

[Haa06] J. de Haan. "How emergence arises". In: *Ecological complexity* 3.4 (2006), pp. 293–301 (Cited on page 75).

[HAC16] HACKADAY. *Alan Yates on the Impossible Task of Making Valve's VR Work*. [Accessed: June 29th, 2023]. 2016. URL: `https://www.youtube.com/watch?v=75ZytcYANTA` (Cited on page 26).

[Hac86] W. Hacker. *Arbeitspsychologie: Psychische Regulation von Arbeitstätigkeiten*. Huber, 1986 (Cited on pages 103, 107).

[HS14] K. S. Hale and K. M. Stanney. *Handbook of virtual environments: Design, implementation, and applications*. CRC Press, 2014 (Cited on pages 27, 33, 35).

[Ham05] B. Hampe. "Image schemas in cognitive linguistics: Introduction". In: *From perception to meaning: Image schemas in cognitive linguistics* 29 (2005), pp. 1–12 (Cited on page 38).

[HMS22] J. Han, A. V. Moere, and A. L. Simeone. "Foldable spaces: An overt redirection approach for natural walking in virtual reality". In: *2022 IEEE Conference on Virtual*

Reality and 3D User Interfaces (VR). IEEE. 2022, pp. 167–175 (Cited on pages 60, 61, 64).

[Han+20] S. Han et al. "MEgATrack: monochrome egocentric articulated hand-tracking for virtual reality". In: *ACM Transactions on Graphics (ToG)* 39.4 (2020), pp. 87–1 (Cited on page 27).

[HWS97] P. Hansson, A. Wallberg, and K. Simsarian. "Techniques for "natural" interaction in multi-user CAVE-like environments". In: *Poster in ECSCW* 97 (1997) (Cited on page 17).

[Har+14] A. Harris, K. Nguyen, P. T. Wilson, M. Jackoski, and B. Williams. "Human joystick: Wii-leaning to translate in large virtual environments". In: *Proceedings of the 13th ACM SIGGRAPH international conference on virtual-reality continuum and its applications in industry*. 2014, pp. 231–234 (Cited on page 98).

[HTS07] S. Harrison, D. Tatar, and P. Sengers. "The Three Paradigms of HCI". In: *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. New York: ACM Press, 2007, p. 18 (Cited on pages 129, 181, 185).

[Har06] S. G. Hart. "NASA-task load index (NASA-TLX); 20 years later". In: *Proceedings of the human factors and ergonomics society annual meeting*. Vol. 50. 9. Sage publications Sage CA: Los Angeles, CA. 2006, pp. 904–908 (Cited on page 230).

[Har+15] T. Hartmann et al. "The spatial presence experience scale (SPES)". In: *Journal of Media Psychology* (2015) (Cited on pages 21, 147, 231).

[HP12] R. Hartson and P. S. Pyla. *The UX Book: Process and guidelines for ensuring a quality user experience*. Elsevier, 2012 (Cited on page 16).

[HBK03] M. Hassenzahl, M. Burmester, and F. Koller. "AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität". In: *Mensch & Computer 2003: Interaktion in Bewegung* (2003), pp. 187–196 (Cited on page 16).

[HDG10] M. Hassenzahl, S. Diefenbach, and A. Göritz. "Needs, affect, and interactive products–Facets of user experience". In: *Interacting with computers* 22.5 (2010), pp. 353–362 (Cited on page 16).

[Hay+15] H. Hayakawa, C. L. Fernando, M. Y. Saraiji, K. Minamizawa, and S. Tachi. "Telexistence drone: Design of a flight telexistence system for immersive aerial sports experience". In: *Proceedings of the 6th Augmented Human International Conference*. 2015, pp. 171–172 (Cited on pages 60, 61).

[Hed19] M. M. Hedblom. "Image Schemas and Concept Invention: Cognitive, Logical, and Linguistic Investigations". PhD thesis. Universität Magdeburg, 2019 (Cited on page 39).

[Hei60] M. L. Heilig. *Stereoscopic-television apparatus for individual use*. U.S. Patent US2955156A. 1960 (Cited on page 23).

[HE21] M. Heras-Escribano. "Pragmatism, enactivism, and ecological psychology: towards a unified approach to post-cognitivism". In: *Synthese* 198.Suppl 1 (2021), pp. 337–363 (Cited on page 49).

[Hew+92] T. T. Hewett, R. Baecker, S. Card, T. Carey, J. Gasen, M. Mantei, G. Perlman, G. Strong, and W. Verplank. *ACM SIGCHI curricula for human-computer interaction*. ACM, 1992 (Cited on page 10).

[Hir+22] L. Hirsch, J. Li, S. Mayer, and A. Butz. "A Survey of Natural Design for Interaction". In: *Proceedings of Mensch und Computer 2022*. 2022, pp. 240–254 (Cited on page 17).

[Hof+20] M. Hofer, T. Hartmann, A. Eden, R. Ratan, and L. Hahn. "The role of plausibility in the experience of spatial presence in virtual environments". In: *Frontiers in Virtual Reality* (2020), p. 2 (Cited on pages 21, 78, 110).

[Hol+10] N. Hollender, C. Hofmann, M. Deneke, and B. Schmitz. "Integrating cognitive load theory and concepts of human–computer interaction". In: *Computers in human behavior* 26.6 (2010), pp. 1278–1288 (Cited on page 42).

[Hol10] D. L. Holton. *Constructivism+ embodied cognition= enactivism: theoretical and practical implications for conceptual change*. 2010 (Cited on page 44).

[HO17] K. Hornbæk and A. Oulasvirta. "What Is Interaction?" In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. CHI '17. Denver, Colorado, USA: Association for Computing Machinery, 2017, pp. 5040–5052 (Cited on page 130).

[HHS19] G. Hovhannisyan, A. Henson, and S. Sood. "Enacting virtual reality: the philosophy and cognitive science of optimal virtual experience". In: *Augmented Cognition: 13th International Conference*. Springer. 2019, pp. 225–255 (Cited on pages 48, 77).

[Hua+21] L. Huang, B. Zhang, Z. Guo, Y. Xiao, Z. Cao, and J. Yuan. "Survey on depth and RGB image-based 3D hand shape and pose estimation". In: *Virtual Reality & Intelligent Hardware* 3.3 (2021), pp. 207–234 (Cited on page 27).

[Hur11] J. Hurtienne. "Image schemas and design for intuitive use". PhD thesis. TU Berlin, 2011 (Cited on pages 38, 39, 94).

[HI07] J. Hurtienne and J. H. Israel. "Image schemas and their metaphorical extensions: intuitive patterns for tangible interaction". In: *Proceedings of the 1st international conference on Tangible and embedded interaction*. 2007, pp. 127–134 (Cited on page 39).

[Hur17] J. Hurtienne. "How Cognitive Linguistics Inspires HCI: Image Schemas and Image-Schematic Metaphors". In: *International Journal of Human–Computer Interaction* 33.1 (2017), pp. 1–20 (Cited on pages 39, 94).

[HL19] M. Husung and E. Langbehn. "Of portals and orbs: An evaluation of scene transition techniques for virtual reality". In: *Proceedings of Mensch Und Computer 2019*. 2019, pp. 245–254 (Cited on pages 60, 61, 82, 84, 86, 117).

[Hut22] D. Hutto. "Getting real about pretense: A radical enactivist proposal". In: *Phenomenology and the Cognitive Sciences* 21.5 (2022), pp. 1157–1175 (Cited on page 48).

[Hva+17a] J. Hvass, O. Larsen, K. Vendelbo, N. Nilsson, R. Nordahl, and S. Serafin. "Visual realism and presence in a virtual reality game". In: *2017 3DTV conference: The true vision-capture, Transmission and Display of 3D video (3DTV-CON)*. IEEE. 2017, pp. 1–4 (Cited on page 61).

[Hva+17b] J. S. Hvass, O. Larsen, K. B. Vendelbo, N. C. Nilsson, R. Nordahl, and S. Serafin. "The effect of geometric realism on presence in a virtual reality game". In: *2017 IEEE Virtual Reality (VR)*. IEEE. 2017, pp. 339–340 (Cited on page 61).

[Ihd90] D. Ihde. "Technology and the lifeworld: From garden to earth". In: (1990) (Cited on pages 73, 74, 94, 99).

[Ihd17] D. Ihde. *Postphenomenology and Technoscience: The Peking University Lectures*. New York, USA: State University of New York Press, 2017 (Cited on pages 72, 73, 179).

[IJs05] W. A. IJsselsteijn. "History of telepresence". In: *3D Videocommunication: Algorithms, Concepts and Real-Time Systems in Human Centred Communication* (2005), pp. 5–21 (Cited on page 19).

[Inc13] P. Inc. *iPhone 1 - Steve Jobs MacWorld keynote in 2007 - Full Presentation, 80 mins*. [Accessed: August 6th, 2023]. 2013. URL: `https://www.youtube.com/watch?v=VQKMoT-6XSg` (Cited on page 53).

[Int18] S. International. *1968 "Mother of All Demos" by SRI's Doug Engelbart and Team*. `https://www.youtube.com/watch?v=B6rKUf9DWRI`. [Accessed: August 17th, 2022]. 2018 (Cited on page 11).

[IRA07] V. Interrante, B. Ries, and L. Anderson. "Seven league boots: A new metaphor for augmented locomotion through moderately large scale immersive virtual environments". In: *2007 IEEE Symposium on 3D User interfaces*. IEEE. 2007 (Cited on pages 29, 97, 111, 117).

[ISO22] ISO25000. *ISO/IEC 25010*. Available at `https://iso25000.com/index.php/en/iso-25000-standards/iso-25010?start=3`. [Accessed: November 17th, 2021]. 2022 (Cited on page 16).

[Isr+15] A. Israr, S. Zhao, K. McIntosh, J. Kang, Z. Schwemler, E. Brockmeyer, M. Baskinger, and M. Mahler. "Po2: augmented haptics for interactive gameplay". In: *ACM SIGGRAPH 2015 Emerging Technologies*. 2015, pp. 1–1 (Cited on page 61).

[Ito+16] Y. Itoh, J. Orlosky, K. Kiyokawa, and G. Klinker. "Laplacian vision: Augmenting motion prediction via optical see-through head-mounted displays". In: *Proceedings of the 7th Augmented Human International Conference 2016*. 2016, pp. 1–8 (Cited on page 112).

[Jac+08a] R. J. 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*. 2008, pp. 201–210 (Cited on pages 22, 53, 111, 116, 123).

[Jac+08b] R. J. 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*. 2008, pp. 201–210 (Cited on page 39).

[JFH94] R. H. Jacoby, M. Ferneau, and J. Humphries. "Gestural interaction in a virtual environment". In: *Stereoscopic Displays and Virtual Reality Systems*. Vol. 2177. SPIE. 1994, pp. 355–364 (Cited on page 35).

[Jer15] J. Jerald. *The VR book: Human-centered design for virtual reality*. Morgan & Claypool, 2015 (Cited on pages 15, 17, 33, 35, 40).

[JRG14] H.-C. Jetter, H. Reiterer, and F. Geyer. "Blended Interaction: understanding natural human–computer interaction in post-WIMP interactive spaces". In: *Personal and Ubiquitous Computing* 18.5 (2014), pp. 1139–1158 (Cited on pages 39, 69, 108).

[Joh65] E. A. Johnson. "Touch display? a novel input/output device for computers". In: *Electronics Letters* 8.1 (1965), pp. 219–220 (Cited on pages 11, 53).

[Joh87] M. Johnson. "The body in the mind: The bodily basis of meaning, imagination, and reason". In: (1987) (Cited on page 38).

[Kai+11] M. Kaipainen, N. Ravaja, P. Tikka, R. Vuori, R. Pugliese, M. Rapino, and T. Takala. "Enactive systems and enactive media: embodied human-machine coupling beyond interfaces". In: *Leonardo* 44.5 (2011), pp. 433–438 (Cited on pages 48, 181).

[Kap71] I. Kapandji. "The physiology of the joints, volume I, upper limb". In: *American Journal of Physical Medicine & Rehabilitation* 50.2 (1971), p. 96 (Cited on page 33).

[Kap18] E. Kapp. *Elements of a philosophy of technology: On the evolutionary history of culture*. U of Minnesota Press, 2018 (Cited on pages 70, 71, 112).

[Kap96] V. Kaptelinin. "Activity theory: Implications for human-computer interaction". In: *Context and consciousness: Activity theory and human-computer interaction* 1.103-116 (1996), p. 1 (Cited on pages 49, 50).

[KN12] V. Kaptelinin and B. Nardi. "Affordances in HCI: Toward a Mediated Action Perspective". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '12. Austin, Texas, USA: Association for Computing Machinery, 2012, pp. 967–976 (Cited on page 127).

[KN97] V. Kaptelinin and B. A. Nardi. "Activity Theory: Basic Concepts and Applications". In: *CHI '97 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '97. Atlanta, Georgia: Association for Computing Machinery, 1997, pp. 158–159 (Cited on pages 49, 127).

[KN06] V. Kaptelinin and B. A. Nardi. *Acting with technology: Activity theory and interaction design*. Cambridge, MA, USA: MIT press, 2006 (Cited on page 49).

[KS05] M. Karam and M. Schraefel. "A taxonomy of gestures in human computer interactions". In: (2005) (Cited on pages 34, 158).

[Ke+20] L. Ke, A. Kamat, J. Wang, T. Bhattacharjee, C. Mavrogiannis, and S. S. Srinivasa. "Telemanipulation with chopsticks: Analyzing human factors in user demonstrations". In: *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2020 (Cited on page 109).

[Ker+23] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis. "3d gaussian splatting for real-time radiance field rendering". In: *ACM Transactions on Graphics (ToG)* 42.4 (2023), pp. 1–14 (Cited on page 152).

[Key+18] C. Keyworth, J. Hart, C. Armitage, and M. P. Tully. "What maximizes the effectiveness and implementation of technology-based interventions to support healthcare professional practice? A systematic literature review". In: *BMC medical informatics and decision making* 18.1 (2018), pp. 1–21 (Cited on page 136).

[KFD15] S. Khan, K. Francis, and B. Davis. "Accumulation of experience in a vast number of cases: Enactivism as a fit framework for the study of spatial reasoning in mathematics education". In: *ZDM* 47 (2015), pp. 269–279 (Cited on pages 41, 42, 107).

[Kic16] Kickstarter. *Oculus Rift: Step Into the Game*. https://www.kickstarter.com/ projects/1523379957/oculus-rift-step-into-the-game/description [Accessed: June 30th, 2023]. 2016 (Cited on page 24).

[KBS13] K. Kilteni, I. Bergstrom, and M. Slater. "Drumming in immersive virtual reality: the body shapes the way we play". In: *IEEE transactions on visualization and computer graphics* 19.4 (2013), pp. 597–605 (Cited on page 127).

[KGS12] K. Kilteni, R. Groten, and M. Slater. "The sense of embodiment in virtual reality". In: *Presence: Teleoperators and Virtual Environments* 21.4 (2012), pp. 373–387 (Cited on pages 20, 81, 82).

[Kim+08] J.-S. Kim, D. Gračanin, K. Matković, and F. Quek. "Finger walking in place (FWIP): A traveling technique in virtual environments". In: *Smart Graphics: 9th International Symposium*. Springer. 2008, pp. 58–69 (Cited on pages 98, 115, 117).

[Kim20] J. Kim. "The problem of nonhuman agency and bodily intentionality in the Anthropocene". In: *Neohelicon* 47.1 (2020), pp. 9–16 (Cited on page 82).

[Kim05] M. Kimmel. "Culture regained: Situated and compound image schemas". In: *From perception to meaning: Image schemas in cognitive linguistics* (2005), pp. 285–311 (Cited on page 37).

[KM94] D. Kirsh and P. Maglio. "On distinguishing epistemic from pragmatic action". In: *Cognitive science* 18.4 (1994), pp. 513–549 (Cited on page 54).

[Kit+99] Y. Kitamura, T. Higashi, T. Masaki, and F. Kishino. "Virtual chopsticks: Object manipulation using multiple exact interactions". In: *Proceedings IEEE Virtual Reality (Cat. No. 99CB36316)*. IEEE. 1999, pp. 198–204 (Cited on page 109).

[Kna+19] L. Knaack, A.-K. Lache, O. Preikszas, S. Reinhold, and M. Teistler. "Improving Readability of Text in Realistic Virtual Reality Scenarios: Visual Magnification Without Restricting User Interactions". In: *Proceedings of Mensch und Computer 2019*. 2019, pp. 749–753 (Cited on page 155).

[KSH19] M. Kocur, V. Schwind, and N. Henze. "Utilizing the Proteus Effect to Improve Interactions using Full-Body Avatars in Virtual Reality". In: *Mensch und Computer 2019 - Workshopband*. Bonn: Gesellschaft für Informatik e.V., 2019 (Cited on page 127).

[KM05] B. Kopp and H. Mandl. *Wissensschemata*. 2005 (Cited on page 39).

[KRR18] O. Koskinen, I. Rakkolainen, and R. Raisamo. "Gigapixel virtual reality employing live superzoom cameras". In: *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*. 2018, pp. 1–2 (Cited on page 155).

[Kos+17] F. Kosmalla, A. Zenner, M. Speicher, F. Daiber, N. Herbig, and A. Krüger. "Exploring rock climbing in mixed reality environments". In: *Proceedings of the 2017 CHI*

Conference Extended Abstracts on Human Factors in Computing Systems. 2017, pp. 1787–1793 (Cited on page 114).

[KS14] S. M. Kosslyn and E. E. Smith. *Cognitive Psychology: Mind and Brain*. Pearson Education Limited, 2014 (Cited on pages 42, 44).

[Kre+21] A. Krekhov, K. Emmerich, R. Rotthaler, and J. Krueger. "Puzzles Unpuzzled: Towards a Unified Taxonomy for Analog and Digital Escape Room Games". In: *Proceedings of the ACM on Human-Computer Interaction* 5.CHI PLAY (2021), pp. 1–24 (Cited on page 61).

[Kri05] K. Krippendorff. *The semantic turn: A new foundation for design*. Boca Raton, FL, USA: CRC Press, 2005 (Cited on pages 103, 106, 107, 135).

[Kri18] K. Krippendorff. *Content analysis: An introduction to its methodology*. Sage publications, 2018 (Cited on page 118).

[Kru+16] D. Krupke, P. Lubos, L. Demski, J. Brinkhoff, G. Weber, F. Willke, and F. Steinicke. "Control methods in a supernatural flight simulator". In: *2016 IEEE Virtual Reality (VR)*. 2016, pp. 329–329 (Cited on pages 60, 61, 64, 76).

[Kuc19] U. Kuckartz. "Qualitative text analysis: A systematic approach". In: *Compendium for early career researchers in mathematics education*. Springer, Cham, 2019, pp. 181–197 (Cited on page 108).

[Kul09] A. Kulik. "Building on realism and magic for designing 3D interaction techniques". In: *IEEE Computer Graphics and Applications* 29.6 (2009), pp. 22–33 (Cited on pages 22, 54, 113).

[KAN16] A. Kultima, K. Alha, and T. Nummenmaa. "Design constraints in game design case: survival mode game jam 2016". In: *Proceedings of the international conference on game jams, hackathons, and game creation events*. 2016, pp. 22–29 (Cited on page 61).

[KL19] K. Kunze and S. Lukosch. "Superhuman sports—a testing ground for augmenting our senses". In: *XRDS: Crossroads, The ACM Magazine for Students* 25.4 (2019), pp. 38–43 (Cited on page 70).

[Kun+17] K. Kunze, K. Minamizawa, S. Lukosch, M. Inami, and J. Rekimoto. "Superhuman sports: Applying human augmentation to physical exercise". In: *IEEE Pervasive Computing* 16.2 (2017), pp. 14–17 (Cited on page 56).

[Kuo+09] L.-C. Kuo, H.-Y. Chiu, C.-W. Chang, H.-Y. Hsu, and Y.-N. Sun. "Functional workspace for precision manipulation between thumb and fingers in normal hands". In: *Journal of electromyography and kinesiology* 19.5 (2009), pp. 829–839 (Cited on page 33).

[KMB94] G. Kurtenbach, T. P. Moran, and W. Buxton. "Contextual animation of gestural commands". In: *Computer Graphics Forum*. Vol. 13. 5. Wiley Online Library. 1994, pp. 305–314 (Cited on page 98).

[Kus15] A. Kush. "Sixth sense technology, a new paradigm". In: *2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions)*. IEEE. 2015, pp. 1–3 (Cited on pages 60, 61).

[Kus14] D. Kushner. "Virtual reality's moment". In: *Ieee Spectrum* 51.1 (2014), pp. 34–37 (Cited on page 2).

[Lab67] A. F. C. R. Laboratories. *Report on Research at AFCRL*. Clearinghouse for Federal Scientific and Technical Information, 1967 (Cited on page 11).

[Lad+19] P. Ladwig, B. Dewitz, H. Preu, and M. Säger. "Remote guidance for machine maintenance supported by physical leds and virtual reality". In: *Kultur und Informatik: Extended Reality*. 2019, pp. 255–262 (Cited on page 221).

[Lae+20] F. Laera, M. M. Foglia, A. Evangelista, A. Boccaccio, M. Gattullo, M Vito, J. L. Gabbard, E Antonio, M. Fiorentino, et al. "Towards sailing supported by augmented reality: Motivation, methodology and perspectives". In: *2020 IEEE International*

Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct). IEEE. 2020, pp. 269–274 (Cited on pages 60, 61).

[Lak12] G. Lakoff. "Explaining embodied cognition results". In: *Topics in cognitive science* 4.4 (2012), pp. 773–785 (Cited on pages 37, 38, 105).

[Lak14] G. Lakoff. "Mapping the brain's metaphor circuitry: metaphorical thought in everyday reason". In: *Frontiers in human neuroscience* 8 (2014), p. 958 (Cited on page 38).

[LLS18] E. Langbehn, P. Lubos, and F. Steinicke. "Evaluation of locomotion techniques for room-scale vr: Joystick, teleportation, and redirected walking". In: *Proceedings of the Virtual Reality International Conference-Laval Virtual*. 2018, pp. 1–9 (Cited on page 106).

[Lan88] J. Lanier. *A Vintage Virtual Reality Interview*. [Accessed: January 26th, 2023]. 1988. URL: `http://www.jaronlanier.com/vrint.html` (Cited on pages 17, 18, 28, 54).

[Lan96] E. Lantz. "The future of virtual reality: head mounted displays versus spatially immersive displays (panel)". In: *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. 1996, pp. 485–486 (Cited on page 125).

[LS87] J. H. Larkin and H. A. Simon. "Why a diagram is (sometimes) worth ten thousand words". In: *Cognitive science* 11.1 (1987), pp. 65–100 (Cited on page 99).

[LW22] M. E. Latoschik and C. Wienrich. "Congruence and plausibility, not presence: Pivotal conditions for XR experiences and effects, a novel approach". In: *Frontiers in Virtual Reality* 3 (2022) (Cited on page 111).

[LJ+17] J. J. LaViola Jr, E. Kruijff, R. P. McMahan, D. Bowman, and I. P. Poupyrev. *3D user interfaces: theory and practice*. Addison-Wesley Professional, 2017 (Cited on pages 10, 15, 18, 22, 24–26, 28–32, 35, 41, 43, 54, 68, 104, 105, 165).

[LBS85] S. Lee, W. Buxton, and K. C. Smith. "A multi-touch three dimensional touch-sensitive tablet". In: *Acm Sigchi Bulletin* 16.4 (1985), pp. 21–25 (Cited on page 53).

[LLK20] A. Lehto, N. Luostarinen, and P. Kostia. "Augmented reality gaming as a tool for subjectivizing visitor experience at cultural heritage locations—case lights on!" In: *Journal on Computing and Cultural Heritage (JOCCH)* 13.4 (2020), pp. 1–16 (Cited on page 61).

[LDH13] E. Lenz, S. Diefenbach, and M. Hassenzahl. "Exploring relationships between interaction attributes and experience". In: *Proceedings of the 6th international conference on designing pleasurable products and interfaces*. 2013, pp. 126–135 (Cited on pages 60, 61).

[Li22] W. Li. "Simulating Ice Skating Experience in Virtual Reality". In: *2022 7th International Conference on Image, Vision and Computing (ICIVC)*. IEEE. 2022, pp. 706–712 (Cited on page 114).

[Lie+22] A. Liebrecht et al. "ARMAGNI: Augmented Reality Enhanced Surgical Magnifying Glasses". In: *Scandinavian Conference on Health Informatics*. 2022, pp. 46–51 (Cited on page 221).

[LR09] S. Lim and B. Reeves. "Being in the game: Effects of avatar choice and point of view on psychophysiological responses during play". In: *Media psychology* 12.4 (2009), pp. 348–370 (Cited on page 127).

[Lin66] N. Lindgren. "Human factors in engineering Part II – Advanced man-machine systems and concepts". In: *IEEE Spectrum* 3.4 (1966), pp. 62–72 (Cited on pages 10, 11).

[Lin+11] A. R. Lingley et al. "A single-pixel wireless contact lens display". In: *Journal of Micromechanics and Microengineering* 21.12 (2011) (Cited on page 183).

[Lin+21] S. Linxen, C. Sturm, F. Brühlmann, V. Cassau, K. Opwis, and K. Reinecke. "How WEIRD is CHI?" In: *Proceedings of the 2021 CHI Conference on Human Factors*

in Computing Systems. CHI '21. Yokohama, Japan: Association for Computing Machinery, 2021 (Cited on page 128).

[Liu+22] P. Liu, E. R. Stepanova, A. Kitson, T. Schiphorst, and B. E. Riecke. "Virtual Transcendent Dream: Empowering People through Embodied Flying in Virtual Reality". In: *CHI Conference on Human Factors in Computing Systems*. 2022, pp. 1–18 (Cited on page 114).

[Liv+13] M. A. Livingston, A. Dey, C. Sandor, and B. H. Thomas. "Pursuit of "X-ray vision" for augmented reality". In: *Human Factors in Augmented Reality Environments*. Springer, 2013, pp. 67–107 (Cited on page 112).

[Lon+22] K. H. Long, K. R. McLellan, M. Boyarinova, and S. J. Bensmaia. "Proprioceptive sensitivity to imposed finger deflections". In: *Journal of Neurophysiology* 127.2 (2022), pp. 412–420 (Cited on page 33).

[LBC04] J. Looser, M. Billinghurst, and A. Cockburn. "Through the looking glass: the use of lenses as an interface tool for Augmented Reality interfaces". In: *Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Australasia and South East Asia*. 2004, pp. 204–211 (Cited on page 155).

[Loo+07] J. Looser, M. Billinghurst, R. Grasset, and A. Cockburn. "An evaluation of virtual lenses for object selection in augmented reality". In: *Proceedings of the 5th international conference on Computer graphics and interactive techniques in Australia and Southeast Asia*. 2007, pp. 203–210 (Cited on page 155).

[LBS14] P. Lubos, G. Bruder, and F. Steinicke. "Are 4 Hands Better than 2? Bimanual Interaction for Quadmanual User Interfaces". In: *Proceedings of the 2nd ACM Symposium on Spatial User Interaction*. SUI '14. Honolulu, Hawaii, USA: Association for Computing Machinery, 2014, pp. 123–126 (Cited on pages 60, 61, 71, 76, 80).

[Lub18] P. B. Lubos. "Supernatural and comfortable user interfaces for basic 3d interaction tasks". PhD thesis. Staats-und Universitätsbibliothek Hamburg Carl von Ossietzky, 2018 (Cited on pages 58–63, 66, 69, 108–110, 114).

[Luc07] A. Luciani. *Virtual reality and virtual environment*. 2007 (Cited on page 18).

[Mac+99] R. C. MacCallum, K. F. Widaman, S. Zhang, and S. Hong. "Sample size in factor analysis." In: *Psychological methods* 4.1 (1999), p. 84 (Cited on page 123).

[Mac+14] S. Mache, K. Vitzthum, B. F. Klapp, and G. Danzer. "Surgeons' work engagement: Influencing factors and relations to job and life satisfaction". In: *The surgeon* 12.4 (2014), pp. 181–190 (Cited on pages 131, 135).

[Mac12] I. S. MacKenzie. "Human-computer interaction: An empirical research perspective". In: (2012) (Cited on page 12).

[Mac18] I. S. MacKenzie. "Fitts' law". In: *Handbook of human-computer interaction* 1 (2018), pp. 349–370 (Cited on page 162).

[MI08] I. S. MacKenzie and P. Isokoski. "Fitts' throughput and the speed-accuracy tradeoff". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2008, pp. 1633–1636 (Cited on pages 160, 171).

[MC14] J. M. Mandler and C. P. Cánovas. "On defining image schemas". In: *Language and cognition* 6.4 (2014), pp. 510–532 (Cited on page 39).

[MM95] D. P. Mapes and J. M. Moshell. "A Two-Handed Interface for Object Manipulation in Virtual Environments". In: *Presence: Teleoperators and Virtual Environments* 4.4 (Nov. 1995), pp. 403–416 (Cited on pages 31, 158).

[Mar93] A. Marcus. "Future directions in advanced user interface design". In: *Communicating with virtual worlds*. Springer. 1993, pp. 2–13 (Cited on page 13).

[Mar98] A. Marcus. "Metaphor design in user interfaces". In: *ACM SIGDOC Asterisk Journal of Computer Documentation* 22.2 (1998), pp. 43–57 (Cited on pages 13, 14).

[Mar+19] D. Mardanbegi, B. Mayer, K. Pfeuffer, S. Jalaliniya, H. Gellersen, and A. Perzl. "EyeSeeThrough: Unifying tool selection and application in virtual environments". In: *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE. 2019, pp. 474–483 (Cited on page 155).

[Mar+18] J. Martinez, D. Griffiths, V. Biscione, O. Georgiou, and T. Carter. "Touchless haptic feedback for supernatural VR experiences". In: *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE. 2018, pp. 629–630 (Cited on pages 60, 61).

[MBM24] G. Marzi, M. Balzano, and D. Marchiori. "K-Alpha Calculator—Krippendorff's Alpha Calculator: A User-Friendly Tool for Computing Krippendorff's Alpha Inter-Rater Reliability Coefficient". In: *MethodsX* (2024) (Cited on page 118).

[Mat10] M. Matsunaga. "How to Factor-Analyze Your Data Right: Do's, Don'ts, and How-To's." In: *International journal of psychological research* 3.1 (2010), pp. 97–110 (Cited on page 116).

[Mat70] H. Maturana. *Biology of cognition*. Biological Computer Laboratory, Department of Electrical Engineering, 1970 (Cited on page 46).

[Mat02] H. Maturana. "Autopoiesis, structural coupling and cognition: a history of these and other notions in the biology of cognition". In: *Cybernetics & human knowing* 9.3-4 (2002), pp. 5–34 (Cited on page 127).

[MV87] H. Maturana and F. Varela. "The tree of life". In: *Shambhala, Boston* (1987) (Cited on page 45).

[Mat+21] I. Matyash, R. Kutzner, T. Neumuth, and M. Rockstroh. "Accuracy measurement of HoloLens2 IMUs in medical environments". In: *Current Directions in Biomedical Engineering* 7.2 (2021), pp. 633–636 (Cited on page 138).

[MB05] A. McMahan and W. Buckland. "Cognitive schemas and virtual reality". In: *Intelligent Agent*. Vol. 5. 2005 (Cited on page 40).

[MLP16] R. P. McMahan, C. Lai, and S. K. Pal. "Interaction fidelity: the uncanny valley of virtual reality interactions". In: *Virtual, Augmented and Mixed Reality: 8th International Conference, VAMR 2016, Held as Part of HCI International 2016, Toronto, Canada, July 17-22, 2016. Proceedings 8*. Springer. 2016, pp. 59–70 (Cited on pages 4, 21, 29, 63, 78, 110, 180, 181).

[McM11] R. P. McMahan. "Exploring the effects of higher-fidelity display and interaction for virtual reality games". PhD thesis. Virginia Tech, 2011 (Cited on pages 22, 111).

[MS+18] J. McVeigh-Schultz, M. Kreminski, K. Prasad, P. Hoberman, and S. S. Fisher. "Immersive design fiction: Using VR to prototype speculative interfaces and interaction rituals within a virtual storyworld". In: *Proceedings of the 2018 designing interactive systems conference*. 2018, pp. 817–829 (Cited on page 61).

[Med+19] D. Medeiros, M. Sousa, A. Raposo, and J. Jorge. "Magic carpet: Interaction fidelity for flying in vr". In: *IEEE transactions on visualization and computer graphics* 26.9 (2019), pp. 2793–2804 (Cited on pages 61, 114).

[Mee+03] M. Meehan, S. Razzaque, M. C. Whitton, and F. P. Brooks. "Effect of latency on presence in stressful virtual environments". In: *IEEE Virtual Reality, 2003. Proceedings.* IEEE. 2003, pp. 141–148 (Cited on page 20).

[MH19] E. D. Mekler and K. Hornbæk. "A framework for the experience of meaning in human-computer interaction". In: *Proceedings of the 2019 CHI conference on human factors in computing systems*. 2019, pp. 1–15 (Cited on page 91).

[MP02] M. Merleau-Ponty. *Phenomenology of perception*. Routledge, 2002 (Cited on page 72).

[Met24] Meta. *The Metaverse and How We'll Build It Together – Connect 2021*. Accessed: February 12th, 2024]. 2024. URL: `https://www.youtube.com/watch?v=Uvufun6xer8` (Cited on page 3).

[Met18] T. K. Metzinger. "Why is virtual reality interesting for philosophers?" In: *Frontiers in Robotics and AI* 5 (2018), p. 101 (Cited on page 84).

[MB22] R. Meyer and N. Brancazio. "Putting down the revolt: Enactivism as a philosophy of nature". In: *Frontiers in Psychology* 13 (2022) (Cited on pages 48, 129).

[Mic22] Microsoft. *MixedReality-WebRTC*. https://github.com/microsoft/MixedReality-WebRTC [Accessed: Feburary 24, 2020]. 2022 (Cited on page 140).

[Mic16] Microsoft HoloLens. *Mixed Reality Blends the Physical and Virtual Worlds*. Accessed: February 12th, 2024]. 2016. URL: `https://www.youtube.com/watch?v=_xpI0JosYUk` (Cited on page 3).

[MK94] P. Milgram and F. Kishino. "A taxonomy of mixed reality visual displays". In: *IEICE TRANSACTIONS on Information and Systems* 77.12 (1994), pp. 1321–1329 (Cited on pages 18, 19).

[Mil17] G. A. Miller. "MAN – COMPUTER INTERACTION". In: *Communication Processes: Proceedings of a Symposium Held in Washington, 1963*. Vol. 4. Elsevier. 2017, p. 228 (Cited on page 10).

[Mil75] P. L. Miller. "An Adaptive Natural Language Parser". In: *American Journal of Computational Linguistics* (1975), pp. 42–56 (Cited on pages 44, 108).

[Min+21] A. Minami, H. Takahashi, Y. Nakata, H. Sumioka, and H. Ishiguro. "The Neighbor in My Left Hand: Development and Evaluation of an Integrative Agent System With Two Different Devices". In: *IEEE Access* 9 (2021) (Cited on page 61).

[MBJS97] M. R. Mine, F. P. Brooks Jr, and C. H. Sequin. "Moving objects in space: exploiting proprioception in virtual-environment interaction". In: *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*. 1997, pp. 19–26 (Cited on page 35).

[Min80] M. Minsky. *Telepresence*. 1980 (Cited on page 19).

[Mir+19] S. Mirhosseini, P. Ghahremani, S. Ojar, J. Marino, and A. Kaufrnan. "Exploration of Large Omnidirectional Images in Immersive Environments". In: *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE. 2019, pp. 413–422 (Cited on page 155).

[Mit+17] R. Mitchell, J. Nishida, E. Encinas, and S. Kasahara. "We-Coupling! Designing New Forms of Embodied Interpersonal Connection". In: *Proceedings of the Eleventh International Conference on Tangible, Embedded, and Embodied Interaction*. 2017, pp. 775–780 (Cited on page 61).

[Moh+06] C. Mohs, J. Hurtienne, J. H. Israel, A. B. Naumann, M. C. Kindsmüller, H. A. Meyer, and A. Pohlmeyer. "IUUI – intuitive use of user interfaces". In: *Tagungsband UP06* (2006) (Cited on page 17).

[MN90] R. Molich and J. Nielsen. "Improving a human-computer dialogue". In: *Communications of the ACM* 33.3 (1990), pp. 338–348 (Cited on page 16).

[Mor18] D. Moran. "What is the phenomenological approach? Revisiting intentional explication". In: *Phenomenology and Mind* 15 (2018), pp. 72–90 (Cited on page 72).

[MSS14] A. E. Mostafa, E. Sharlin, and M. C. Sousa. "Poster: Superhumans: A 3DUI design metaphor". In: *2014 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE. 2014, pp. 143–144 (Cited on pages 61, 63).

[Mun+02] A. Munro, R. Breaux, J. Patrey, and B. Sheldon. "Cognitive aspects of virtual environments design". In: *Handbook of virtual environments*. CRC Press, 2002, pp. 455–474 (Cited on page 104).

[MJ62] B. B. Murdock Jr. "The serial position effect of free recall." In: *Journal of experimental psychology* 64.5 (1962), p. 482 (Cited on page 44).

[Mwa01] D. Mwanza. "Where Theory meets Practice: A Case for an Activity Theory based Methodology to guide Computer System Design". In: *Proceedings of INTERACT'*

2001: Eighth IFIP TC 13 Conference on Human-Computer Interaction. Oxford: IOS Press Oxford, 2001 (Cited on page 50).

[Myt+10] O. T. Mytton et al. "Introducing new technology safely". In: *BMJ Quality & Safety* 19.Suppl 2 (2010), pp. i9–i14 (Cited on pages 134, 136).

[NB15] M. Nabioyuni and D. A. Bowman. "An evaluation of the effects of hyper-natural components of interaction fidelity on locomotion performance in virtual reality". In: *Proceedings of the 25th International Conference on Artificial Reality and Telexistence and 20th Eurographics Symposium on Virtual Environments*. 2015, pp. 167–174 (Cited on pages 56, 58, 59, 62, 63, 66, 67, 110, 114–116, 123).

[Nab15] M. Nabiyouni. "How Does Interaction Fidelity Influence User Experience in VR Locomotion?" PhD thesis. Virginia Tech, 2015 (Cited on pages 58–63).

[Nab+15] M. Nabiyouni, A. Saktheeswaran, D. A. Bowman, and A. Karanth. "Comparing the performance of natural, semi-natural, and non-natural locomotion techniques in virtual reality". In: *2015 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE. 2015, pp. 3–10 (Cited on pages 29, 112).

[Nag80] T. Nagel. "What is it like to be a bat?" In: *The language and thought series*. Harvard University Press, 1980, pp. 159–168 (Cited on page 79).

[NS19] R. Nakagawa and K. Sonobe. "Encounters: A multiparticipant audiovisual art experience with XR". In: *SIGGRAPH Asia 2019 XR*. 2019, pp. 6–8 (Cited on pages 60, 61).

[Nar97] S. Narayanan. "Talking the talk is like walking the walk: A computational model of verbal aspect". In: *Proceedings of the 19th Cognitive Science Society Conference*. Citeseer. 1997, pp. 548–553 (Cited on pages 38, 94).

[Nar98] B. Nardi. *Context and consciousness: Activity theory and human-computer interaction*. 1998 (Cited on pages 49, 71).

[Nar96] B. A. Nardi. "Studying context: A comparison of activity theory, situated action models, and distributed cognition". In: *Context and consciousness: Activity theory and human-computer interaction* (1996), pp. 35–52 (Cited on page 49).

[Nel+82] T. O. Nelson, J. Leonesio, A. P. Shimamura, R. F. Landwehr, and L. Narens. "Overlearning and the feeling of knowing." In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 8.4 (1982), p. 279 (Cited on pages 99, 103, 105).

[NC85] A. Newell and S. K. Card. "The prospects for psychological science in human-computer interaction". In: *Human-computer interaction* 1.3 (1985), pp. 209–242 (Cited on page 125).

[New91] K. M. Newell. "Motor skill acquisition". In: *Annual review of psychology* 42.1 (1991), pp. 213–237 (Cited on pages 36, 42).

[New18] A. Newen. "The embodied self, the pattern theory of self, and the predictive mind". In: *Frontiers in psychology* 9 (2018), p. 2270 (Cited on pages 74–76, 83).

[New+22] M. Newman, B Gatersleben, K. Wyles, and E Ratcliffe. "The use of virtual reality in environment experiences and the importance of realism". In: *Journal of environmental psychology* 79 (2022), p. 101733 (Cited on page 21).

[NCL17] A. K. T. Ng, L. K. Y. Chan, and H. Y. K. Lau. "A low-cost lighthouse-based virtual reality head tracking system". In: *2017 International Conference on 3D Immersion (IC3D)*. 2017, pp. 1–5 (Cited on page 26).

[Ngu14] T. T. H. Nguyen. "Proposition of new metaphors and techniques for 3D interaction and navigation preserving immersion and facilitating collaboration between distant users". PhD thesis. INSA de Rennes, 2014 (Cited on pages 58–61, 64).

[Nie86] J. Nielsen. "A virtual protocol model for computer-human interaction". In: *International Journal of Man-Machine Studies* 24.3 (1986), pp. 301–312 (Cited on page 11).

[Nie94] J. Nielsen. *Usability engineering*. San Fransisco: Morgan Kaufmann, 1994 (Cited on pages 16, 28).

[NM90] J. Nielsen and R. Molich. "Heuristic evaluation of user interfaces". In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. 1990, pp. 249–256 (Cited on page 16).

[NNS16] N. Nilsson, R. Nordahl, and S. Serafin. "Immersion revisited: A review of existing definitions of immersion and their relation to different theories of presence". In: *Human technology* 12.2 (2016), p. 108 (Cited on page 19).

[Nil15] N. C. Nilsson. "Walking Without Moving: An exploration of factors influencing the perceived naturalness of Walking-in-Place techniques for locomotion in virtual environments." PhD thesis. Aalborg Universitetsforlag. Ph.d.-serien for Det Teknisk-Naturvidenskabelige Fakultet, Aalborg Universitet, 2015 (Cited on pages 64, 111).

[NSN16] N. C. Nilsson, S. Serafin, and R. Nordahl. "Walking in place through virtual worlds". In: *International Conference on Human-Computer Interaction*. Springer. 2016, pp. 37–48 (Cited on pages 30, 55, 66, 123).

[Nil+18] N. C. Nilsson, S. Serafin, F. Steinicke, and R. Nordahl. "Natural walking in virtual reality: A review". In: *Computers in Entertainment (CIE)* 16.2 (2018), pp. 1–22 (Cited on page 114).

[Noë04] A. Noë. *Action in perception*. Cambridge, MA, USA: MIT press, 2004 (Cited on pages 47, 48, 80).

[Nor13] D. Norman. *The design of everyday things: Revised and expanded edition*. New York: Basic books, 2013 (Cited on pages 12, 14, 15).

[Nor88] D. A. Norman. *The psychology of everyday things*. Basic books, 1988 (Cited on pages 103–107).

[Nor98] D. A. Norman. *The invisible computer: why good products can fail, the personal computer is so complex, and information appliances are the solution*. Cambridge, MA, USA: MIT press, 1998 (Cited on pages 11, 105, 125, 126).

[Nor10] D. A. Norman. "Natural user interfaces are not natural". In: *interactions* 17.3 (2010), pp. 6–10 (Cited on pages 12, 17, 108).

[NS86] D. A. Norman and T. Shallice. "Attention to action: Willed and automatic control of behavior". In: *Consciousness and self-regulation: Advances in research and theory volume 4*. Springer, 1986, pp. 1–18 (Cited on pages 36, 104, 105).

[Obe+18] M. Oberhauser, D. Dreyer, R. Braunstingl, and I. Koglbauer. "What's real about virtual reality flight simulation? Comparing the fidelity of a virtual reality with a conventional flight simulation environment." In: *Aviation Psychology and Applied Human Factors* 8.1 (2018), p. 22 (Cited on page 114).

[Oes81] R. Oesterreich. *Handlungsregulation und Kontrolle*. Urban & Schwarzenberg, 1981 (Cited on pages 103, 106).

[O'h+13] K. O'hara, R. Harper, H. Mentis, A. Sellen, and A. Taylor. "On the naturalness of touchless: putting the "interaction" back into NUI". In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 20.1 (2013), pp. 1–25 (Cited on page 16).

[Olv+22] C. Olvera, G. Lara, A. Valdivia, and A. Peña. "A Literature Review of Hand-Based Interaction in Virtual Environments Through Computer Vision". In: *New Perspectives in Software Engineering: Proceedings of the 10th International Conference on Software Process Improvement (CIMPS 2021) 10*. Springer. 2022, pp. 113–122 (Cited on pages 27, 34).

[Ort+16] F. R. Ortega, F. Abyarjoo, A. Barreto, N. Rishe, and M. Adjouadi. *Interaction design for 3D user interfaces: The world of modern input devices for research, applications, and game development*. CRC Press, 2016 (Cited on page 16).

[OI04] D. O'Sullivan and T. Igoe. *Physical computing: sensing and controlling the physical world with computers*. Boston, MA, USA: Course Technology Press, 2004 (Cited on pages 79, 80, 127).

[Pag+21] M. J. Page et al. "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews". In: *BMJ* (Mar. 2021) (Cited on pages 57, 58).

[Pan80] A. Pantages. "Oral history of captain grace hopper". In: *Computer History Museum* (1980), p. 20 (Cited on page 10).

[PBC21] S. Park, S. Bokijonov, and Y. Choi. "Review of microsoft hololens applications over the past five years". In: *Applied Sciences* 11.16 (2021), p. 7259 (Cited on page 131).

[Pau91] R. Pausch. "Virtual reality on five dollars a day". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1991, pp. 265–270 (Cited on page 23).

[Pau+96] R. Pausch, J. Snoddy, R. Taylor, S. Watson, and E. Haseltine. "Disney's Aladdin: first steps toward storytelling in virtual reality". In: *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. 1996, pp. 193–203 (Cited on pages 109, 114).

[PB37] W. Penfield and E. Boldrey. "Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation". In: *Brain* 60.4 (1937), pp. 389–443 (Cited on pages 42, 43, 79).

[PR50] W. Penfield and T. Rasmussen. *The cerebral cortex of man; a clinical study of localization of function.* Macmillan, 1950 (Cited on pages 42, 43).

[Pfe+17] K. Pfeuffer, B. Mayer, D. Mardanbegi, and H. Gellersen. "Gaze+ pinch interaction in virtual reality". In: *Proceedings of the 5th symposium on spatial user interaction*. 2017, pp. 99–108 (Cited on pages 60, 61).

[Pia52] J. Piaget. *The origins of intelligence in children*. International Universities Press, 1952 (Cited on pages 37, 103, 107).

[Pic+21] T. Picker, B. Dewitz, C. Geiger, and F. Steinicke. "Echtzeit-Fingertracking in Unity 3D durch 3D Convolutional Neural Networks in einem Multi-Depth-Camera-Setup". In: *GI VR/AR Workshop*. Gesellschaft für Informatik eV. 2021 (Cited on page 221).

[PDG18] T. Picker, B. Dewitz, and C. Geiger. "Fingertracking durch neuronale Netze anhand reduzierter Markersets und Motion-Capture-Daten". In: (2018) (Cited on page 221).

[Pie+97] J. S. Pierce, A. S. Forsberg, M. J. Conway, S. Hong, R. C. Zeleznik, and M. R. Mine. "Image plane interaction techniques in 3D immersive environments". In: *Proceedings of the 1997 symposium on Interactive 3D graphics*. 1997, 39–ff (Cited on page 155).

[Pit17] F. Pittarello. "Experimenting with PlayVR, a virtual reality experience for the world of theater". In: *Proceedings of the 12th biannual conference on Italian SIGCHI chapter*. 2017, pp. 1–10 (Cited on pages 60, 61, 64).

[Pit+21] D. Pittera, O. Georgiou, A. Abdouni, and W. Frier. ""I Can Feel It Coming in the Hairs Tonight": Characterising Mid-Air Haptics on the Hairy Parts of the Skin". In: *IEEE Transactions on Haptics* 15.1 (2021), pp. 188–199 (Cited on page 61).

[Pla23] Playstation. *PS VR2 Headset Teardown Video - First Look with Engineers Behind the Next-Gen Hardware*. https://www.youtube.com/watch?v=NYhngu66Ccc [Accessed: June 30th, 2023]. 2023 (Cited on page 26).

[PD13] S. Poeschl and N. Doering. "The German VR Simulation Realism Scale–psychometric construction for virtual reality applications with virtual humans". In: *Annual Review of Cybertherapy and Telemedicine 2013* (2013), pp. 33–37 (Cited on page 21).

[PRL20] Y. B. Popova and J. Rączaszek-Leonardi. "Enactivism and ecological psychology: The role of bodily experience in agency". In: *Frontiers in Psychology* 11 (2020), p. 539841 (Cited on page 127).

[PT22] L. Poretski and A. Tang. "Press A to Jump: Design Strategies for Video Game Learnability". In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 2022, pp. 1–26 (Cited on pages 60, 61).

[Pou+96a] I. Poupyrev, M. Billinghurst, S. Weghorst, and T. Ichikawa. "The go-go interaction technique: non-linear mapping for direct manipulation in VR". In: *Proceedings of the 9th annual ACM symposium on User interface software and technology*. 1996, pp. 79–80 (Cited on pages 31, 71, 75, 76, 84, 97).

[Pou+96b] I. Poupyrev, M. Billinghurst, S. Weghorst, and T. Ichikawa. "The go-go interaction technique: non-linear mapping for direct manipulation in VR". In: *Proceedings of the 9th annual ACM symposium on User interface software and technology*. 1996, pp. 79–80 (Cited on page 70).

[PSR15] J. Preece, H. Sharp, and Y. Rogers. *Interaction design: beyond human-computer interaction*. John Wiley & Sons, 2015 (Cited on page 16).

[PBU93] J. Preece, D. Benyon, and O. University. *A guide to usability: Human factors in computing*. Addison-Wesley Longman Publishing Co., Inc., 1993 (Cited on page 16).

[PMW22] L. M. Prinz, T. Mathew, and B. Weyers. "A Systematic Literature Review of Virtual Reality Locomotion Taxonomies". In: *IEEE transactions on visualization and computer graphics* (2022) (Cited on pages 60, 61, 64).

[Pri97] W. Prinz. "Perception and action planning". In: *European journal of cognitive psychology* 9.2 (1997), pp. 129–154 (Cited on page 105).

[Que04] W. Quesenbery. "Balancing the 5Es of usability". In: *Cutter IT Journal* 17.2 (2004), pp. 4–11 (Cited on page 16).

[RMC15] S. Radmard, A. J. Moon, and E. A. Croft. "Interface design and usability analysis for a robotic telepresence platform". In: *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE. 2015, pp. 511–516 (Cited on page 171).

[Rag+15] E. D. Ragan, D. A. Bowman, R. Kopper, C. Stinson, S. Scerbo, and R. P. McMahan. "Effects of Field of View and Visual Complexity on Virtual Reality Training Effectiveness for a Visual Scanning Task". In: *IEEE Transactions on Visualization and Computer Graphics* 21.7 (2015), pp. 794–807 (Cited on pages 21, 22).

[Rag+20] K. Ragozin, D. Zheng, G. Chernyshov, and D. Hynds. "Sophroneo: Fear not. a VR horror game with thermal feedback and physiological signal loop". In: *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 2020, pp. 1–6 (Cited on page 61).

[Rak+20] I. Rakkolainen, E. Freeman, A. Sand, R. Raisamo, and S. Brewster. "A survey of mid-air ultrasound haptics and its applications". In: *IEEE Transactions on Haptics* 14.1 (2020), pp. 2–19 (Cited on pages 61, 166).

[Ran+23] H.-R. Rantamaa, J. Kangas, S. K. Kumar, H. Mehtonen, J. Järnstedt, and R. Raisamo. "Comparison of a VR Stylus with a Controller, Hand Tracking, and a Mouse for Object Manipulation and Medical Marking Tasks in Virtual Reality". In: *Applied Sciences* 13.4 (2023), p. 2251 (Cited on page 171).

[Ras83] J. Rasmussen. "Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models". In: *IEEE transactions on systems, man, and cybernetics* 3 (1983), pp. 257–266 (Cited on page 103).

[Ras87] J. Rasmussen. *Mental models and the control of actions in complex environments*. Vol. 59. Risø National Laboratory, 1987 (Cited on pages 14, 15, 104, 106).

[Rau+22] P. A. Rauschnabel, R. Felix, C. Hinsch, H. Shahab, and F. Alt. "What is XR? Towards a framework for augmented and virtual reality". In: *Computers in Human Behavior* 133 (2022), p. 107289 (Cited on page 19).

[RA12] G. M. Rayan and E. Akelman. *The hand: anatomy, examination, and diagnosis*. Lippincott Williams & Wilkins, 2012 (Cited on page 33).

[Ray09] C. Raymaekers. "Special issue on enactive interfaces". In: *Interacting with Computers* 21.1-2 (Jan. 2009), pp. 1–2. ISSN: 0953-5438. DOI: 10.1016/j.intcom.2008.10.010. eprint: https://academic.oup.com/iwc/article-pdf/21/1-2/1/2153500/iwc21-0001.pdf. URL: https://doi.org/10.1016/j.intcom.2008.10.010 (Cited on page 48).

[Raz05] S. Razzaque. *Redirected walking*. The University of North Carolina at Chapel Hill, 2005 (Cited on pages 29, 114).

[RS20] C. Read and A. Szokolszky. "Ecological psychology and enactivism: perceptually-guided action vs. sensation-based enaction". In: *Frontiers in psychology* 11 (2020), p. 1270 (Cited on pages 46, 49).

[REN89] B. RENO. "Full field of view dome display system". In: *Flight Simulation Technologies Conference and Exhibit*. 1989, p. 3316 (Cited on page 24).

[Ric+15] S. Richir, P. Fuchs, D. Lourdeaux, D. Millet, C. Buche, and R. Querrec. "How to design compelling Virtual Reality or Augmented Reality experience?" In: *International Journal of Virtual Reality* 15.1 (2015), pp. 35–47 (Cited on page 40).

[RZ21] B. E. Riecke and D. Zielasko. "Continuous vs. Discontinuous (Teleport) Locomotion in VR: How Implications can Provide both Benefits and Disadvantages". In: *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 2021, pp. 373–374. DOI: 10.1109/VRW52623.2021.00075 (Cited on pages 60, 61, 64).

[RTO19] F. E. Ritter, F. Tehranchi, and J. D. Oury. "ACT-R: A cognitive architecture for modeling cognition". In: *Wiley Interdisciplinary Reviews: Cognitive Science* 10.3 (2019) (Cited on page 103).

[Rob+86] D. W. Roberts, J. W. Strohbehn, J. F. Hatch, W. Murray, and H. Kettenberger. "A frameless stereotaxic integration of computerized tomographic imaging and the operating microscope". In: *Journal of Neurosurgery* 65.4 (1986), pp. 545–549 (Cited on page 130).

[Rog+22] K. Rogers, S. Karaosmanoglu, M. Altmeyer, A. Suarez, and L. E. Nacke. "Much Realistic, Such Wow! A Systematic Literature Review of Realism in Digital Games". In: *CHI Conference on Human Factors in Computing Systems*. 2022, pp. 1–21 (Cited on pages 20, 182).

[Roh95] T. Rohrer. "Feelings Stuck in a GUI Web: Metaphors, Image-Schemata, and Designing the Human-Computer Interface". In: *Center for the Cognitive Science of Metaphor, Philosophy Department, University of Oregon.* (1995) (Cited on page 39).

[RM+20a] E. Rojas-Muñoz, M. E. Cabrera, C. Lin, D. Andersen, V. Popescu, K. Anderson, B. L. Zarzaur, B. Mullis, and J. P. Wachs. "The System for Telementoring with Augmented Reality (STAR): A head-mounted display to improve surgical coaching and confidence in remote areas". In: *Surgery* 167.4 (2020), pp. 724–731 (Cited on pages 130, 183).

[RM+20b] E. Rojas-Muñoz et al. "Evaluation of an augmented reality platform for austere surgical telementoring: a randomized controlled crossover study in cricothyroidotomies". In: *NPJ digital medicine* 3.1 (2020), pp. 1–9 (Cited on pages 130, 140).

[RVF22] G. Rolla, G. Vasconcelos, and N. M. Figueiredo. "Virtual reality, embodiment, and allusion: An ecological-enactive approach". In: *Philosophy & Technology* 35.4 (2022), p. 95 (Cited on pages 48, 88).

[RL+78] E. Rosch, B. B. Lloyd, et al. "Cognition and categorization". In: (1978) (Cited on pages 38, 94).

[RBB13] R. S. Rosenberg, S. L. Baughman, and J. N. Bailenson. "Virtual superheroes: Using superpowers in virtual reality to encourage prosocial behavior". In: *PloS one* 8.1 (2013), e55003 (Cited on pages 56, 76).

[Ros17] R. Rosenberger. "Notes on a nonfoundational phenomenology of technology". In: *Foundations of Science* 22 (2017), pp. 471–494 (Cited on page 73).

[RV15] R. Rosenberger and P. P. Verbeek. *Postphenomenological investigations: Essays on human-technology relations*. Blue Ridge Summit: Lexington Books, 2015 (Cited on pages 72–74, 85, 86, 99).

[RHG03] J. C. Rosser, B. Herman, and L. E. Giammaria. "Telementoring". In: 10.4 (2003), pp. 209–217 (Cited on page 130).

[RYK07] J. C. Rosser, S. M. Young, and J. Klonsky. "Telementoring: An application whose time has come". In: *Surgical Endoscopy and Other Interventional Techniques* 21.8 (2007), pp. 1458–1463 (Cited on page 130).

[ROS08] M. Roussou, M. Oliver, and M. Slater. "Exploring activity theory as a tool for evaluating interactivity and learning in virtual environments for children". In: *Cognition, Technology & Work* 10 (2008), pp. 141–153 (Cited on page 50).

[Rou+20] F.-E. Roux, M. Niare, S. Charni, C. Giussani, and J.-B. Durand. "Functional architecture of the motor homunculus detected by electrostimulation". In: *The Journal of Physiology* 598.23 (2020) (Cited on page 42).

[Row11] M. J. Rowlands. *Body language: Representation in action*. MIT Press, 2011 (Cited on page 89).

[Rub46] S. L. Rubinshtein. *Foundations of general psychology*. 1946 (Cited on page 49).

[Rum17] D. E. Rumelhart. "Schemata: The building blocks of cognition". In: *Theoretical issues in reading comprehension*. Routledge, 2017, pp. 33–58 (Cited on pages 36, 96, 100, 107).

[RY18] D. M. Russell and S. Yarosh. "Can we look to science fiction for innovation in HCI?" In: *Interactions* 25.2 (2018), pp. 36–40 (Cited on page 67).

[SH21a] S. Sadeghian and M. Hassenzahl. "From Limitations to "Superpowers": A Design Approach to Better Focus on the Possibilities of Virtual Reality to Augment Human Capabilities". In: *Designing Interactive Systems Conference 2021*. 2021, pp. 180–189 (Cited on pages 55, 56, 112).

[San02] E. B.-N. Sanders. "From user-centered to participatory design approaches". In: *Design and the social sciences*. Boca Raton, FL, USA: CRC Press, 2002, pp. 18–25 (Cited on page 126).

[SCL17] B. Sarupuri, M. L. Chipana, and R. W. Lindeman. "Trigger walking: A low-fatigue travel technique for immersive virtual reality". In: *2017 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE. 2017, pp. 227–228 (Cited on page 115).

[Sau+22] Y. Sauer, A. Sipatchin, S. Wahl, and M. García García. "Assessment of consumer VR-headsets' objective and subjective field of view (FoV) and its feasibility for visual field testing". In: *Virtual Reality* 26.3 (2022), pp. 1089–1101 (Cited on page 28).

[SSH20] E. Sayyad, M. Sra, and T. Höllerer. "Walking and teleportation in wide-area virtual reality experiences". In: *2020 IEEE international symposium on mixed and augmented reality (ISMAR)*. IEEE. 2020, pp. 608–617 (Cited on pages 60, 61, 64).

[SES99] D. Schmalstieg, L. M. Encarnação, and Z. Szalavári. "Using transparent props for interaction with the virtual table". In: *Proceedings of the 1999 symposium on Interactive 3D graphics*. 1999, pp. 147–153 (Cited on page 155).

[SS99] D. Schmalstieg and G. Schaufler. "Sewing worlds together with SEAMS: A mechanism to construct complex virtual environments". In: *Presence* 8.4 (1999), pp. 449–461 (Cited on page 61).

[SBTP20] T. Schofield, J. Bowers, and D. Trujillo Pisanty. "Magical Realist Design". In: *Proceedings of the 2020 ACM Designing Interactive Systems Conference*. 2020, pp. 1873–1886 (Cited on pages 60, 61).

[Sch+95] L. Schomaker et al. "A taxonomy of multimodal interaction in the human information processing system". In: (1995) (Cited on page 40).

[Sch08] R. Schroeder. “Defining virtual worlds and virtual environments”. In: *Journal For Virtual Worlds Research* 1.1 (2008) (Cited on page 19).

[SFR01] T. Schubert, F. Friedmann, and H. Regenbrecht. “The experience of presence: Factor analytic insights”. In: *Presence: Teleoperators & Virtual Environments* 10.3 (2001), pp. 266–281 (Cited on page 21).

[Sch+19a] P. Schulz, D. Alexandrovsky, F. Putze, R. Malaka, and J. Schöning. “The role of physical props in vr climbing environments”. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 2019, pp. 1–13 (Cited on page 114).

[ST55] R. J. Schwarz and C. Taylor. “The anatomy and mechanics of the human hand”. In: *Artificial limbs* 2.2 (1955), pp. 22–35 (Cited on page 33).

[Sch+19b] 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*. 2019, pp. 1–12 (Cited on page 20).

[Ser+18] M. Serpi, A. Carcangiu, A. Murru, and L. D. Spano. “Web5VR: a flexible framework for integrating virtual reality input and output devices on the web”. In: *Proceedings of the ACM on Human-Computer Interaction* 2.EICS (2018), pp. 1–19 (Cited on pages 54, 68).

[SPR19] H. Sharp, J. Preece, and Y. Rogers. *Interaction Design: Beyond Human-Computer Interaction*. 2019 (Cited on page 14).

[Shl12] A. Shleifer. “Psychologists at the gate: a review of Daniel Kahneman’s thinking, fast and slow”. In: *Journal of Economic Literature* 50.4 (2012), pp. 1080–1091 (Cited on page 41).

[Shn97] B. Shneiderman. “Direct manipulation for comprehensible, predictable and controllable user interfaces”. In: *Proceedings of the 2nd international conference on Intelligent user interfaces*. 1997, pp. 33–39 (Cited on page 12).

[Shn03] B. Shneiderman. “Why not make interfaces better than 3D reality?” In: *IEEE Computer Graphics and Applications* 23.6 (2003), pp. 12–15 (Cited on pages 54, 68).

[Sim+22] A. L. Simeone, R. Cools, S. Depuydt, J. M. Gomes, P. Goris, J. Grocott, A. Esteves, and K. Gerling. “Immersive speculative enactments: bringing future scenarios and technology to life using virtual reality”. In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 2022, pp. 1–20 (Cited on page 61).

[SBW17] R. Skarbez, F. P. Brooks Jr, and M. C. Whitton. “A survey of presence and related concepts”. In: *ACM Computing Surveys (CSUR)* 50.6 (2017), pp. 1–39 (Cited on pages 19–21).

[SSW21] R. Skarbez, M. Smith, and M. C. Whitton. “Revisiting milgram and kishino’s reality-virtuality continuum”. In: *Frontiers in Virtual Reality* 2 (2021), p. 647997 (Cited on page 18).

[Sla09] M. Slater. “Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 364.1535 (2009), pp. 3549–3557 (Cited on pages 19, 79, 94, 135).

[SU93] M. Slater and M. Usoh. “Presence in immersive virtual environments”. In: *Proceedings of IEEE virtual reality annual international symposium*. IEEE. 1993, pp. 90–96 (Cited on page 21).

[SU94] M. Slater and M. Usoh. “Body centred interaction in immersive virtual environments”. In: *Artificial life and virtual reality* 1.1994 (1994), pp. 125–148 (Cited on pages 55, 66, 68, 113).

[SW97] M. Slater and S. Wilbur. “A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments”. In: *Presence: Teleoperators & Virtual Environments* 6.6 (1997), pp. 603–616 (Cited on pages 18, 20).

[Sme+17] J. Smeddinck, D. Alexandrovsky, D. Wenig, M. Zimmer, W. Wegele, S. Juergens, and R. Malaka. “Hoverboard: A Leap to the Future of Locomotion in VR!?” In:

International Conference on Entertainment Computing. Springer. 2017, pp. 218–225 (Cited on page 114).

[Smi86] R. B. Smith. "Experiences with the alternate reality kit: an example of the tension between literalism and magic". In: *ACM SIGCHI Bulletin* 17.SI (1986), pp. 61–67 (Cited on pages 53, 56, 68, 69).

[Soa+05] L. P. Soares, L. Nomura, M. C. Cabral, M. Nagamura, R. de Deus Lopes, and M. K. Zuffo. "Virtual hang-gliding over Rio de Janeiro." In: *SIGGRAPH Emerging Technologies*. 2005, p. 29 (Cited on page 114).

[SHP18] B. Son, C. Hunihan, and S. Prakkamakul. "SoundGlove: Multisensory Exploration of Everyday Objects for Creative Purposes". In: *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. 2018, pp. 1–6 (Cited on pages 60, 61).

[Son+12] P. Song, W. B. Goh, W. Hutama, C.-W. Fu, and X. Liu. "A Handle Bar Metaphor for Virtual Object Manipulation with Mid-Air Interaction". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '12. Austin, Texas, USA: Association for Computing Machinery, 2012, pp. 1297–1306 (Cited on page 158).

[SFS12] K. Sonntag, E. Frieling, and R. Stegmaier. *Lehrbuch Arbeitspsychologie*. Hogrefe AG, 2012 (Cited on page 107).

[SR03] C. R. de Souza and D. F. Redmiles. *Using activity theory to understand contradictions in collaborative software development*. 2003 (Cited on page 50).

[Spe+18a] M. Speicher, A. M. Feit, P. Ziegler, and A. Krüger. "Selection-based text entry in virtual reality". In: *Proceedings of the 2018 CHI conference on human factors in computing systems*. 2018, pp. 1–13 (Cited on page 61).

[Spe+18b] M. Speicher, P. Hell, F. Daiber, A. Simeone, and A. Krüger. "A virtual reality shopping experience using the apartment metaphor". In: *Proceedings of the 2018 International Conference on Advanced Visual Interfaces*. 2018, pp. 1–9 (Cited on pages 60, 61, 64).

[Spi51] H. Spiegelberg. "Supernaturalism or naturalism: A study in meaning and verifiability". In: *Philosophy of Science* 18.4 (1951), pp. 339–368 (Cited on page 68).

[Ste14] P. Steiner. *Enacting Anti-Representationalism. The Scope and the Limits of Enactive Critiques of Representationalism*. 2014. DOI: 10.26913/50202014.0109.0003 (Cited on pages 126, 129).

[SL10] M. Steinert and L. Leifer. "Scrutinizing Gartner's hype cycle approach". In: *Picmet 2010 technology management for global economic growth*. IEEE. 2010, pp. 1–13 (Cited on page 3).

[Ste16a] F. Steinicke. *Being really virtual*. Springer, 2016 (Cited on pages 2, 24, 183).

[Ste16b] F. Steinicke. *Frank Steinicke (Professor, Uni. Hamburg): Super-Natural User Interfaces for the Ultimate Display*. Talk at Augmented World Expo, available at https://www.youtube.com/watch?v=OiHVDp_foio. [Accessed: July 5th, 2022]. 2016 (Cited on pages 58, 59, 62, 63, 110).

[Ste17a] F. Steinicke. "Fooling your senses: (super-) natural user interfaces for the ultimate display". In: *Proceedings of the 5th Symposium on Spatial User Interaction*. 2017, pp. 1–2 (Cited on pages 58, 61, 62, 66, 114).

[Ste17b] F. Steinicke. *Fooling your Senses: (Super-)Natural User Interfaces for the Ultimate Display*. Keynote at SUI '17: ACM Symposium on Spatial User Interaction, available at https://www.youtube.com/watch?v=5xQu3XZwVuQ. [Accessed: July 5th, 2022]. 2017 (Cited on pages 58, 59, 62, 66, 69, 108).

[SBH07] F. Steinicke, G. Bruder, and K. Hinrichs. "Hybrid traveling in fully-immersive large-scale geographic environments". In: *Proceedings of the 2007 ACM symposium on Virtual reality software and technology*. 2007, pp. 229–230 (Cited on page 114).

[Ste+09] F. Steinicke, G. Bruder, J. Jerald, H. Frenz, and M. Lappe. "Estimation of detection thresholds for redirected walking techniques". In: *IEEE transactions on visualization and computer graphics* 16.1 (2009), pp. 17–27 (Cited on pages 29, 111, 114).

[SW13] A. Stephan and S. Walter. *Handbuch Kognitionswissenschaft*. Heidelberg, Germany: Springer, 2013 (Cited on page 48).

[SH21b] P. Stilwell and K. Harman. "Phenomenological research needs to be renewed: Time to integrate enactivism as a flexible resource". In: *International Journal of Qualitative Methods* 20 (2021), p. 1609406921995299 (Cited on pages 48, 72).

[SCP95] R. Stoakley, M. J. Conway, and R. Pausch. "Virtual reality on a WIM: interactive worlds in miniature". In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. 1995, pp. 265–272 (Cited on pages 30, 35, 155).

[Sto18] T. A. Stoffregen. "Affordances as properties of the animal-environment system". In: *How Shall Affordances Be Refined?* Routledge, 2018, pp. 115–134 (Cited on page 82).

[SBM06] T. A. Stoffregen, B. G. Bardy, and B. Mantel. "Affordances in the design of enactive systems". In: *Virtual Reality* 10 (2006), pp. 4–10 (Cited on page 48).

[SD04] F. Strack and R. Deutsch. "Reflective and impulsive determinants of social behavior". In: *Personality and social psychology review* 8.3 (2004), pp. 220–247 (Cited on pages 101, 106).

[Str96] G. M. Stratton. "Some preliminary experiments on vision without inversion of the retinal image." In: *Psychological review* 3.6 (1896), p. 611 (Cited on page 23).

[Str99] G. Strawson. "The self and the SESMET". In: *Journal of Consciousness Studies* 6 (1999), pp. 99–135 (Cited on page 74).

[Stu98] S. Stuntz. *MUI Homepage*. `http://www.sasg.com/mui/`. [Accessed: August 29th, 2022]. 1998 (Cited on page 53).

[SZP89] D. J. Sturman, D. Zeltzer, and S. Pieper. "Hands-on interaction with virtual environments". In: *Proceedings of the 2nd annual ACM SIGGRAPH symposium on User interface software and technology*. 1989, pp. 19–24 (Cited on pages 34, 35).

[Stu92] D. J. Sturman. "Whole-hand input". PhD thesis. Massachusetts Institute of Technology, 1992 (Cited on pages 32–34).

[Sut64] I. E. Sutherland. "Sketchpad a man-machine graphical communication system". In: *Simulation* 2.5 (1964), R–3 (Cited on page 11).

[Sut65] I. E. Sutherland. "The ultimate display". In: *Multimedia: From Wagner to virtual reality* 1 (1965) (Cited on pages 2, 3, 18, 20, 54).

[Sut68] I. E. Sutherland. "A head-mounted three dimensional display". In: *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*. 1968, pp. 757–764 (Cited on pages 11, 23).

[Swe03] J. Sweller. "Evolution of human cognitive architecture". In: *Psychology of learning and motivation* 43 (2003), pp. 216–266 (Cited on pages 36–38, 41, 42, 99, 108, 179).

[Sza19] B. K. Szabó. "Interaction in an immersive virtual reality application". In: *2019 10th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*. IEEE. 2019, pp. 35–40 (Cited on page 108).

[TJ01] V. Tanriverdi and R. J. Jacob. "VRID: A Design Model and Methodology for Developing Virtual Reality Interfaces". In: *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*. VRST '01. Baniff, Alberta, Canada: Association for Computing Machinery, 2001, pp. 175–182 (Cited on pages 55, 66).

[TI+21] E. M. Taranta II, C. R. Pittman, M. Maghoumi, M. Maslych, Y. M. Moolenaar, and J. J. Laviola Jr. "Machete: Easy, Efficient, and Precise Continuous Custom Gesture Segmentation". In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 28.1 (2021), pp. 1–46 (Cited on page 61).

[TS21] F. J. Thiel and A. Steed. "" Lend Me a Hand"–Extending the Reach of Seated VR Players in Unmodified Games Through Remote Co-Piloting". In: *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE. 2021, pp. 214–219 (Cited on pages 60, 61).

[Tho07] E. Thompson. *Mind in life: Biology, phenomenology, and the sciences of mind*. Harvard University Press, 2007 (Cited on pages 45, 46).

[Tho16] J. Thomson. *Collected papers in physics and engineering*. Cambridge University Press, 2016 (Cited on page 10).

[TM94] R. A. Thurman and J. S. Mattoon. "Virtual reality: Toward fundamental improvements in simulation-based training". In: *Educational technology* 34.8 (1994), pp. 56–64 (Cited on page 21).

[Tia+18] M. Tiator, C. Geiger, B. Dewitz, B. Fischer, L. Gerhardt, D. Nowottnik, and H. Preu. "Venga! climbing in mixed reality". In: *Proceedings of the First Superhuman Sports Design Challenge: First International Symposium on Amplifying Capabilities and Competing in Mixed Realities*. 2018, pp. 1–8 (Cited on page 221).

[Tom+14] C. Tominski, S. Gladisch, U. Kister, R. Dachselt, and H. Schumann. "A Survey on Interactive Lenses in Visualization." In: *EuroVis (STARs)*. Citeseer. 2014 (Cited on page 155).

[Tre+19] A. Treskunov, E. Gerhardt, D. Nowottnik, B. Fischer, L. Gerhardt, M. Säger, and C. Geiger. "ICAROSmuIti-A VR Test Environment for the Development of Multimodal and Multi-User Interaction Concepts". In: *Proceedings of Mensch und Computer 2019*. 2019, pp. 909–911 (Cited on page 114).

[Tri+20] C. Triebus, B. Dewitz, I. Družetić, C. Huhn, P. Kretschel, and C. Geiger. "is a rose— A Performative Installation in the Context of Art and Technology". In: *Kultur und Informatik: Extended Reality*. 2020, pp. 153–165 (Cited on pages 115, 221).

[Tri+21] C. Triebus, I. Druzetic, B. Dewitz, C. Huhn, P. Kretschel, and C. Geiger. "is a rose– A Performative Installation between the Tangible and the Digital". In: *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction*. 2021, pp. 1–4 (Cited on page 221).

[Tuc04] A. B. Tucker. *Computer science handbook*. Chapman and Hall/CRC, 2004 (Cited on page 13).

[Uso+99] M. Usoh, K. Arthur, M. C. Whitton, R. Bastos, A. Steed, M. Slater, and F. P. Brooks Jr. "Walking> walking-in-place> flying, in virtual environments". In: *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*. 1999, pp. 359–364 (Cited on pages 20, 29).

[VTR93] F. Varela, E. Thompson, and E Rosch. *The Embodied Mind: Cognitive Science and Human Experience (Cambridge, MA and London*. 1993 (Cited on pages 45, 48, 90).

[VB08] V. Venkatesh and H. Bala. "Technology acceptance model 3 and a research agenda on interventions". In: *Decision sciences* 39.2 (2008), pp. 273–315 (Cited on pages 125, 126, 129).

[Ver16] P.-P. Verbeek. "Toward a Theory of Technological Mediation: A Program for Postphenomenological Research". In: *Technoscience and postphenomenology: The Manhattan papers*. Ed. by J. B. O. Friis and R. C. Crease. London: Lexington Books, 2016, pp. 189–204 (Cited on page 72).

[Ver01] P. P. Verbeek. "Don Ihde: The Technological Lifeworld." In: *American Philosophy of Technology: The Empirical Turn.* Indiana University Press, 2001, pp. 119–146 (Cited on pages 72, 94).

[VHS94] H. Veron, P. Hezel, and D. A. Southard. "Head-mounted displays for virtual reality". In: *Helmet-and Head-Mounted Displays and Symbology Design Requirements*. Vol. 2218. SPIE. 1994, pp. 41–50 (Cited on pages 23, 24).

[Vid13] R. V. V. Vidal. "To be human is to be creative". In: *AI & society* 28.2 (2013), pp. 237–248 (Cited on page 115).

[VP21] M. Villalobos and S. Palacios. "Autopoietic theory, enactivism, and their incommensurable marks of the cognitive". In: *Synthese* 198.Suppl 1 (2021), pp. 71–87 (Cited on page 129).

[Vla+21] J. Vlasblom, R. Arents, R. van Gimst, and A. de Reus. "Virtual Cockpit: Making Natural Interaction Possible in a Low-Cost VR Simulator". In: (2021) (Cited on page 114).

[VS17] M. Vosmeer and B. Schouten. "Project Orpheus a research study into 360 cinematic VR". In: *Proceedings of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video*. 2017, pp. 85–90 (Cited on page 61).

[VRC23] VRCompare. *Oculus Quest 2 vs Samsung Odyssey vs Microsoft HoloLens 2 vs HTC Vive Pro Eye (Comparison)*. https://vr-compare.com /compare?h1=pDTZ02PkT&h2=nJ1nfaxkWVN&h3=EkSDYv0cW&h4=gc___iL-JZfK [Accessed: July 1st, 2023]. 2023 (Cited on page 28).

[VC78] L. S. Vygotsky and M. Cole. *Mind in society: Development of higher psychological processes*. Cambridge, MA, USA: Harvard university press, 1978 (Cited on page 49).

[WS20] J. B. Wagman and T. A. Stoffregen. "It doesn't add up: Nested affordances for reaching are perceived as a complex particular". In: *Attention, Perception, & Psychophysics* 82 (2020), pp. 3832–3841 (Cited on pages 82, 95, 179).

[Wal90] R. Walser. "Doing it directly: the experiential design of cyberspaces". In: *Stereoscopic Displays and Applications*. Vol. 1256. SPIE. 1990, pp. 147–153 (Cited on pages 23, 54).

[WHGA20] M. Walther-Hansen and M. Grimshaw-Aagaard. "Don't extend! reduce! the sound approach to reality". In: *Proceedings of the 15th International Conference on Audio Mostly*. 2020, pp. 8–15 (Cited on page 83).

[Wan+19] L. Wang, H. Zhao, Z. Wang, J. Wu, B. Li, Z. He, and V. Popescu. "Occlusion management in vr: A comparative study". In: *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE. 2019, pp. 708–716 (Cited on pages 60, 61, 64).

[War16] D. Ward. "Achieving transparency: An argument for enactivism". In: *Philosophy and Phenomenological Research* 93.3 (2016), pp. 650–680 (Cited on page 129).

[WSV17] D. Ward, D. Silverman, and M. Villalobos. "Introduction: The varieties of enactivism". In: *Topoi* 36 (2017), pp. 365–375 (Cited on pages 46, 48, 129).

[WAB93] C. Ware, K. Arthur, and K. S. Booth. "Fish tank virtual reality". In: *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*. 1993, pp. 37–42 (Cited on page 24).

[Wat68] W. C. Watt. "Habitability". In: *American Documentation* 19.3 (1968), pp. 338–351 (Cited on page 17).

[Wei16] S. G. Weinbaum. *Pygmalion's spectacles*. Simon and Schuster, 2016 (Cited on page 2).

[Wei+12] I. B. Weiner, R. M. Lerner, M. A. Easterbrooks, and J. Mistry. *Handbook of psychology, developmental psychology*. Vol. 6. John Wiley & Sons, 2012 (Cited on pages 102–105).

[WB96] M. Weiser and J. S. Brown. "Designing calm technology". In: *PowerGrid Journal* 1.1 (1996), pp. 75–85 (Cited on pages 15, 81).

[Wex95] A. Wexelblat. "An approach to natural gesture in virtual environments". In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 2.3 (1995), pp. 179–200 (Cited on page 34).

[Whe38] C. Wheatstone. "XVIII. Contributions to the physiology of vision. – Part the first. On some remarkable, and hitherto unobserved, phenomena of binocular vision". In: *Philosophical transactions of the Royal Society of London* 128 (1838), pp. 371–394 (Cited on page 23).

[Whe14] M. Wheeler. *The revolution will not be optimised: enactivism, embodiment and relationality*. 2014 (Cited on page 129).

[Whi+18] D. Whitman, J Love, G Rainville, and L Skufca. "US Public Opinion & Interest on Human Enhancements Technology". In: *Washington, DC: AARP Research. doi: https://doi. org/10.26419/res* 192 (2018) (Cited on page 112).

[WW11] D. Wigdor and D. Wixon. *Brave NUI world: designing natural user interfaces for touch and gesture*. Elsevier, 2011 (Cited on pages 10, 14, 17, 89, 108).

[Wil+21] W. Willett, B. A. Aseniero, S. Carpendale, P. Dragicevic, Y. Jansen, L. Oehlberg, and P. Isenberg. "Perception! Immersion! Empowerment! Superpowers as Inspiration for Visualization". In: *IEEE Transactions on Visualization and Computer Graphics* 28.1 (2021), pp. 22–32 (Cited on pages 54–56, 112).

[WHB06] C. A. Wingrave, Y. Haciahmetoglu, and D. A. Bowman. "Overcoming world in miniature limitations by a scaled and scrolling WIM". In: *3D User Interfaces (3DUI'06)*. IEEE. 2006, pp. 11–16 (Cited on page 155).

[WF86] T. Winograd and F. Flores. *Understanding computers and cognition: A new foundation for design*. Intellect Books, 1986 (Cited on pages 40, 49, 99).

[WMW09] J. O. Wobbrock, M. R. Morris, and A. D. Wilson. "User-defined gestures for surface computing". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM. 2009, pp. 1083–1092 (Cited on page 34).

[WP20] P. Worth and C. Proctor. "Congruence/Incongruence (Rogers)". In: *Encyclopedia of personality and individual differences* (2020), pp. 838–840 (Cited on page 84).

[Woź+21] M. P. Woźniak, P. Sikorski, M. Wróbel-Lachowska, N. Bartłomiejczyk, J. Dominiak, K. Grudzień, and A. Romanowski. "Enhancing in-game immersion using BCI-controlled mechanics". In: *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology*. 2021, pp. 1–6 (Cited on pages 60, 61).

[WP17] M.-L. Wu and V. Popescu. "Efficient VR and AR navigation through multiperspective occlusion management". In: *IEEE transactions on visualization and computer graphics* 24.12 (2017), pp. 3069–3080 (Cited on pages 60, 61, 64).

[Xia+19] J. Xiao, J. Liu, J. Han, and Y. Wang. "Design of achromatic surface microstructure for near-eye display with diffractive waveguide". In: *Optics Communications* 452 (2019), pp. 411–416 (Cited on page 25).

[Yan+19] L. Yang, J. Huang, T. Feng, W. Hong-An, and D. Guo-Zhong. "Gesture interaction in virtual reality". In: *Virtual Reality & Intelligent Hardware* 1.1 (2019), pp. 84–112 (Cited on page 27).

[Yar67] A. L. Yarbus. "Eye movements during perception of complex objects". In: *Eye movements and vision* (1967), pp. 171–211 (Cited on page 44).

[YKT11] I. Yavrucuk, E. Kubali, and O. Tarimci. "A low cost flight simulator using virtual reality tools". In: *IEEE Aerospace and Electronic Systems Magazine* 26.4 (2011), pp. 10–14 (Cited on page 114).

[YLO17] J. Ylipulli, A. Luusua, and T. Ojala. "On Creative Metaphors in Technology Design: Case" Magic"". In: *Proceedings of the 8th International Conference on Communities and Technologies*. 2017, pp. 280–289 (Cited on pages 60, 61).

[Yoo+17] J. W. Yoon, R. E. Chen, K. ReFaey, R. J. Diaz, R. Reimer, R. J. Komotar, A. Quinones-Hinojosa, B. L. Brown, and R. E. Wharen. "Technical feasibility and safety of image-guided parieto-occipital ventricular catheter placement with the assistance of a wearable head-up display". In: *The International Journal of Medical Robotics and Computer Assisted Surgery* 13.4 (2017), e1836 (Cited on page 152).

[You+96] C. Youngblut, R. E. Johnston, S. H. Nash, R. A. Wienclaw, and C. A. Will. "Review of Virtual Environment Interface Technology." In: (1996) (Cited on page 23).

[Yu+22] K. Yu, U. Eck, F. Pankratz, M. Lazarovici, D. Wilhelm, and N. Navab. "Duplicated reality for co-located augmented reality collaboration". In: *IEEE Transactions on Visualization and Computer Graphics* 28.5 (2022), pp. 2190–2200 (Cited on pages 61, 64, 167).

[ZF18] H. Zacher and M. Frese. "Action regulation theory: Foundations, current knowledge and future directions". In: *The SAGE handbook of industrial, work & organizational psychology: Organizational psychology* 2 (2018), pp. 122–144 (Cited on pages 103, 106, 107).

[Zel92] D. Zeltzer. "Autonomy, interaction, and presence". In: *Presence: Teleoperators & Virtual Environments* 1.1 (1992), pp. 127–132 (Cited on pages 18, 20, 23).

[Zha+20] J. Zhang, Z. Dong, R. Lindeman, and T. Piumsomboon. "Spatial scale perception for design tasks in virtual reality". In: *Proceedings of the 2020 ACM Symposium on Spatial User Interaction*. 2020, pp. 1–3 (Cited on page 61).

[Zha00] Z. Zhang. "A flexible new technique for camera calibration". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22.11 (2000), pp. 1330–1334. DOI: 10.1109/34.888718 (Cited on page 142).

[Zha21] Z. Zhang. *Representational Enactivism*. 2021 (Cited on page 129).

[Zie03] T. Ziemke. "What's that thing called embodiment?" In: *Proceedings of the 25th Annual Cognitive Science Society*. Psychology Press, 2003, pp. 1305–1310 (Cited on pages 81, 179).

APPENDICES

A Author Contribution

In the following table A.1, the estimated contribution to the publications listed in section 1.5 is presented following a system proposed by Clement that estimates the author contribution by dividing the responsibilities in a research publication into the four categories *idea*, *work*, *writing*, *stewardship* [Cle14]. In this HCI-focused work, these general aspects are adapted as follows:

→ *Ideas*: Initial conception of the work, developing the research question, designing the experiment.
→ *Work*: Implementation and development, setup, data acquisition, data analysis, framework construction.
→ *Writing*: Drafting the article, producing text, creation of media content, revisiting, approving the final version.
→ *Stewardship*: Acquiring resources for conducting the work, correspondence with publisher, integrity of the work before and after publication.

In the case of stewardship, the following values are estimated: 0% - 20%: minor impact on stewardship and funding was provided for the work, 20 - 50%: shared impact on the integrity and funding was provided for the work, 66%: primary responsibility for the research and no funding provided. For *ideas*, *work*, and *writing*, the values are estimated percentages of work that were confirmed by the main authors of the particular publication (however, in most cases, after several years and not immediately after finalizing the publication). With this model, the overall contribution is then calculated as $\text{Contribution} = 0.2 \cdot \text{Ideas} + 0.3 \cdot \text{Work} + 0.35 \cdot \text{Writing} + 0.15 \cdot \text{Stewardship}$ [Cle14].

Reference	Short Title	Ideas	Work	Writing	Stewards.	Contr.
[DGS23]	Enacting VR Interaction	90%	90%	95%	66%	→ **88%**
[Dew+23b]	Enacted Selves	90%	90%	90%	50%	→ **84%**
[Dew+23a]	Magic or Empowerment?	60%	70%	70%	33%	→ **62%**
[DGS22]	Acting Beyond Reality	90%	90%	95%	66%	→ **88%**
[Dew+22a]	5G Telementoring	33%	75%	80%	33%	→ **62%**
[Dew+22b]	Real-Time Surgery Scans	33%	75%	75%	33%	→ **60%**
[Dew+21]	Zoom Mechanisms	90%	80%	90%	33%	→ **78%**
[DGS21]	Virtual Visus	90%	80%	90%	50%	→ **81%**
[DGS20]	Millimeter Hand Interaction	80%	90%	90%	50%	→ **82%**
[DSG19]	Functional Workspace	80%	80%	90%	50%	→ **76%**
[Dew+18]	REN-Model	80%	80%	90%	33%	→ **76%**
[Lie+22]	ARMAGNI	10%	10%	5%	5%	→ **8%**
[Tri+21]	is a rose II	20%	60%	10%	20%	→ **29%**
[Pic+21]	Finger Tracking II	30%	30%	5%	20%	→ **20%**
[Tri+20]	is a rose I	20%	60%	15%	20%	→ **30%**
[Lad+19]	Remote Guidiance	10%	10%	5%	5%	→ **8%**
[Tia+18]	Venga!	10%	10%	0%	10%	→ **7%**
[Gie+18]	Walking on Walls	40%	30%	5%	33%	→ **24%**
[PDG18]	Finger Tracking I	30%	20%	5%	20%	→ **17%**

Table A.1: Estimated author contribution for main-authored (top) and co-authored (bottom) publications that are related to this dissertation.

B Literature Review Corpus

The following sources were assessed in the literature review (presented in section 3.2):

1 Ando, R., Ando, A., Kunze, K., & Minamizawa, K. (2018). Bubble jumper: enhancing the traditional japanese sport sumo with physical augmentation. In Proceedings of the First Superhuman Sports Design Challenge: First International Symposium on Amplifying Capabilities and Competing in Mixed Realities (pp. 1–6).

2 Apostolopoulos, J., Chou, P., Culbertson, B., Kalker, T., Trott, M., & Wee, S. (2012). The road to immersive communication. Proceedings of the IEEE, 100(4), 974–990.

3 Becker, J., Meyer, U., Eichler, T., & Draheim, S. (2019). A supernatural VR environment for spatial user rotation. In 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR) (pp. 850–851).

4 Blumenthal, H., & Xu, Y. (2012). The ghost club storyscape: designing for transmedia storytelling. IEEE Transactions on Consumer Electronics, 58(2), 190–196.

5 Borütecene, A., & Buruk, O. (2019). Otherworld: Ouija Board as a Resource for Design. In Proceedings of the Halfway to the Future Symposium 2019 (pp. 1–4).

6 Buckers, T., Gong, B., Eisemann, E., & Lukosch, S. (2018). VRabl: stimulating physical activities through a multiplayer augmented reality sports game. In Proceedings of the First Superhuman Sports Design Challenge: First International Symposium on Amplifying Capabilities and Competing in Mixed Realities (pp. 1–5).

7 Byrne, D., Lockton, D., Hu, M., Luong, M., Ranade, A., Escarcha, K., Giesa, K., Huang, Y., Yochum, C., Robertson, G., & others (2022). Spooky Technology: The ethereal and otherworldly as a resource for design. In Designing Interactive Systems Conference (pp. 759–775).

8 Byrne, R., Marshall, J., & Mueller, F. (2016). Balance ninja: towards the design of digital vertigo games via galvanic vestibular stimulation. In Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play (pp. 159–170).

9 Byrne, R., Marshall, J., & Mueller, F. (2016). Designing the vertigo experience: Vertigo as a design resource for digital bodily play. In Proceedings of the TEI'16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction (pp. 296–303).

10 Chittaro, L., & Scagnetto, I. (2001). Is semitransparency useful for navigating virtual environments?. In Proceedings of the ACM symposium on Virtual reality software and technology (pp. 159–166).

11 Cools, R., & Simeone, A. (2021). Mobile Displays for Cross-Reality Interactions between Virtual and Physical Realities. In Proceedings of the 20th International Conference on Mobile and Ubiquitous Multimedia (pp. 217–219).

12 Costa, W., Ananias, L., Barbosa, I., Barbosa, B., De'Carli, A., Barioni, R., Figueiredo, L., Teichrieb, V., & Filgueira, D. (2019). Songverse: a music-loop authoring tool based on Virtual Reality. In 2019 21st Symposium on Virtual and Augmented Reality (SVR) (pp. 216–222).

13 Eghtebas, C., Weber, S., & Klinker, G. (2018). Investigation into Natural Gestures Using EMG for" SuperNatural" Interaction in VR. In Adjunct Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology (pp. 102–104).

14 Fisher, J., Garg, A., Singh, K., & Wang, W. (2017). Designing intentional impossible spaces in virtual reality narratives: A case study. In 2017 IEEE Virtual Reality (VR) (pp. 379–380).

15 Furumoto, T., Ito, M., Fujiwara, M., Makino, Y., Shinoda, H., & Kamigaki, T. (2019). Three-dimensional interaction technique using an acoustically manipulated balloon. Journal is required!, 51–52.

16 Gauglitz, S., Nuernberger, B., Turk, M., & Höllerer, T. (2014). World-stabilized annotations and virtual scene navigation for remote collaboration. In Proceedings of the 27th annual ACM symposium on User interface software and technology (pp. 449–459).

17 Gruenefeld, U., Brück, Y., & Boll, S. (2020). Behind the scenes: Comparing x-ray visualization techniques in head-mounted optical see-through augmented reality. In Proceedings of the 19th International Conference on Mobile and Ubiquitous Multimedia (pp. 179–185).

18 Gugenheimer, J., Dobbelstein, D., Winkler, C., Haas, G., & Rukzio, E. (2016). 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 (pp. 49–60).

19 Han, J., Moere, A., & Simeone, A. (2022). Foldable spaces: An overt redirection approach for natural walking in virtual reality. In 2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR) (pp. 167–175).

20 Hayakawa, H., Fernando, C., Saraiji, M., Minamizawa, K., & Tachi, S. (2015). Telexistence drone: Design of a flight telexistence system for immersive aerial sports experience. In Proceedings of the 6th Augmented Human International Conference (pp. 171–172).

21 Husung, M., & Langbehn, E. (2019). Of portals and orbs: An evaluation of scene transition techniques for virtual reality. Journal is required!, 245–254.

22 Hvass, J., Larsen, O., Vendelbo, K., Nilsson, N., Nordahl, R., & Serafin, S. (2017). The effect of geometric realism on presence in a virtual reality game. In 2017 IEEE Virtual Reality (VR) (pp. 339–340).

23 Hvass, J., Larsen, O., Vendelbo, K., Nilsson, N., Nordahl, R., & Serafin, S. (2017). Visual realism and presence in a virtual reality game. In 2017 3DTV conference: The true vision-capture, Transmission and Display of 3D video (3DTV-CON) (pp. 1–4).

24 Israr, A., Zhao, S., McIntosh, K., Kang, J., Schwemler, Z., Brockmeyer, E., Baskinger, M., & Mahler, M. (2015). Po2: augmented haptics for interactive gameplay. Journal is required!, 1–1.

25 Krekhov, A., Emmerich, K., Rotthaler, R., & Krueger, J. (2021). Puzzles Unpuzzled: Towards a Unified Taxonomy for Analog and Digital Escape Room Games. Proceedings of the ACM on Human-Computer Interaction, 5(CHI PLAY), 1–24.

26 Krupke, D., Lubos, P., Demski, L., Brinkhoff, J., Weber, G., Willke, F., & Steinicke, F. (2016). Control methods in a supernatural flight simulator. In 2016 IEEE Virtual Reality (VR) (pp. 329-329).

27 Kultima, A., Alha, K., & Nummenmaa, T. (2016). Design constraints in game design case: survival mode game jam 2016. In Proceedings of the international conference on game jams, hackathons, and game creation events (pp. 22–29).

28 Kush, A. (2015). Sixth sense technology, a new paradigm. In 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions) (pp. 1–3).

29 Laera, F., Foglia, M., Evangelista, A., Boccaccio, A., Gattullo, M., Vito, M., Gabbard, J., Antonio, E., Fiorentino, M., & others (2020). Towards sailing supported by augmented reality: Motivation, methodology and perspectives. In 2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct) (pp. 269–274).

30 Lehto, A., Luostarinen, N., & Kostia, P. (2020). Augmented reality gaming as a tool for subjectivizing visitor experience at cultural heritage locations—case lights on!. Journal on Computing and Cultural Heritage (JOCCH), 13(4), 1–16.

31 Lenz, E., Diefenbach, S., & Hassenzahl, M. (2013). Exploring relationships between interaction attributes and experience. In Proceedings of the 6th international conference on designing pleasurable products and interfaces (pp. 126–135).

32 Lubos, P. B. (2018). Supernatural and comfortable user interfaces for basic 3d interaction tasks (Doctoral dissertation, Staats-und Universitätsbibliothek Hamburg Carl von Ossietzky).

33 Lubos, P., Bruder, G., & Steinicke, F. (2014, October). Are 4 hands better than 2? bimanual interaction for quadmanual user interfaces. In Proceedings of the 2nd ACM symposium on Spatial user interaction (pp. 123-126).

34 Martinez, J., Griffiths, D., Biscione, V., Georgiou, O., & Carter, T. (2018). Touchless haptic feedback for supernatural VR experiences. In 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR) (pp. 629–630).

35 McVeigh-Schultz, J., Kreminski, M., Prasad, K., Hoberman, P., & Fisher, S. (2018). Immersive design fiction: Using VR to prototype speculative interfaces and interaction rituals within a virtual storyworld. In Proceedings of the 2018 designing interactive systems conference (pp. 817–829).

36 Medeiros, D., Sousa, M., Raposo, A., & Jorge, J. (2019). Magic carpet: Interaction fidelity for flying in vr. IEEE transactions on visualization and computer graphics, 26(9), 2793–2804.

37 Minami, A., Takahashi, H., Nakata, Y., Sumioka, H., & Ishiguro, H. (2021). The Neighbor in My Left Hand: Development and Evaluation of an Integrative Agent System With Two Different Devices. IEEE Access, 9, 98317–98326.

38 Mitchell, R., Nishida, J., Encinas, E., & Kasahara, S. (2017). We-Coupling! Designing New Forms of Embodied Interpersonal Connection. In Proceedings of the Eleventh International Conference on Tangible, Embedded, and Embodied Interaction (pp. 775–780).

39 Mostafa, A., Sharlin, E., & Sousa, M. (2014). Poster: Superhumans: A 3DUI design metaphor. In 2014 IEEE Symposium on 3D User Interfaces (3DUI) (pp. 143–144).

40 Nabiyouni, M., & Bowman, D. (2015). An Evaluation of the Effects of Hyper-Natural Components of Interaction Fidelity on Locomotion Performance in Virtual Reality. In ICAT-EGVE 2015 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments. The Eurographics Association.

41 Nakagawa, R., & Sonobe, K. (2019). Encounters: A multiparticipant audiovisual art experience with XR. Journal is required!, 6–8.

42 Nguyen, T. T. H. (2014). Proposition of new metaphors and techniques for 3D interaction and navigation preserving immersion and facilitating collaboration between distant users (Doctoral dissertation, INSA de Rennes).

43 Pfeuffer, K., Mayer, B., Mardanbegi, D., & Gellersen, H. (2017). Gaze+ pinch interaction in virtual reality. In Proceedings of the 5th symposium on spatial user interaction (pp. 99–108).

44 Pittarello, F. (2017). Experimenting with PlayVR, a virtual reality experience for the world of theater. In Proceedings of the 12th biannual conference on Italian SIGCHI chapter (pp. 1–10).

45 Pittera, D., Georgiou, O., Abdouni, A., & Frier, W. (2021). "I Can Feel It Coming in the Hairs Tonight": Characterising Mid-Air Haptics on the Hairy Parts of the Skin. IEEE Transactions on Haptics, 15(1), 188–199.

46 Poretski, L., & Tang, A. (2022). Press A to Jump: Design Strategies for Video Game Learnability. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (pp. 1–26).

47 Prinz, L., Mathew, T., & Weyers, B. (2022). A Systematic Literature Review of Virtual Reality Locomotion Taxonomies. IEEE transactions on visualization and computer graphics.

48 Ragozin, K., Zheng, D., Chernyshov, G., & Hynds, D. (2020). Sophroneo: Fear not. a VR horror game with thermal feedback and physiological signal loop. In Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1–6).

49 Rakkolainen, I., Freeman, E., Sand, A., Raisamo, R., & Brewster, S. (2020). A survey of mid-air ultrasound haptics and its applications. IEEE Transactions on Haptics, 14(1), 2–19.

50 Riecke, B., & Zielasko, D. (2021). Continuous vs. Discontinuous (Teleport) Locomotion in VR: How Implications can Provide both Benefits and Disadvantages. In 2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW) (pp. 373-374).

51 Sayyad, E., Sra, M., & Höllerer, T. (2020). Walking and teleportation in wide-area virtual reality experiences. In 2020 IEEE international symposium on mixed and augmented reality (ISMAR) (pp. 608–617).

52 Schmalstieg, D., & Schaufler, G. (1999). Sewing worlds together with SEAMS: A mechanism to construct complex virtual environments. Presence, 8(4), 449–461.

53 Schofield, T., Bowers, J., & Trujillo Pisanty, D. (2020). Magical Realist Design. In Proceedings of the 2020 ACM Designing Interactive Systems Conference (pp. 1873–1886).

54 Simeone, A., Cools, R., Depuydt, S., Gomes, J., Goris, P., Grocott, J., Esteves, A., & Gerling, K. (2022). Immersive speculative enactments: bringing future scenarios and technology to life using virtual reality. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (pp. 1–20).

55 Son, B., Hunihan, C., & Prakkamakul, S. (2018). SoundGlove: Multisensory Exploration of Everyday Objects for Creative Purposes. In Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (pp. 1–6).

56 Speicher, M., Feit, A., Ziegler, P., & Krüger, A. (2018). Selection-based text entry in virtual reality. In Proceedings of the 2018 CHI conference on human factors in computing systems (pp. 1–13).

57 Speicher, M., Hell, P., Daiber, F., Simeone, A., & Krüger, A. (2018). A virtual reality shopping experience using the apartment metaphor. In Proceedings of the 2018 International Conference on Advanced Visual Interfaces (pp. 1–9).

58 Steinicke, F. (2017). Fooling your senses: (super-) natural user interfaces for the ultimate display. In Proceedings of the 5th Symposium on Spatial User Interaction (pp. 1–2).

59 Taranta II, E., Pittman, C., Maghoumi, M., Maslych, M., Moolenaar, Y., & Laviola Jr, J. (2021). Machete: Easy, Efficient, and Precise Continuous Custom Gesture Segmentation. ACM Transactions on Computer-Human Interaction (TOCHI), 28(1), 1–46.

60 Thiel, F., & Steed, A. (2021). " Lend Me a Hand"–Extending the Reach of Seated VR Players in Unmodified Games Through Remote Co-Piloting. In 2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW) (pp. 214–219).

61 Vosmeer, M., & Schouten, B. (2017). Project Orpheus a research study into 360 cinematic VR. In Proceedings of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video (pp. 85–90).

62 Wang, L., Zhao, H., Wang, Z., Wu, J., Li, B., He, Z., & Popescu, V. (2019). Occlusion management in vr: A comparative study. In 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR) (pp. 708–716).

63 Woźniak, M., Sikorski, P., Wróbel-Lachowska, M., Bartlomiejczyk, N., Dominiak, J., Grudzień, K., & Romanowski, A. (2021). Enhancing in-game immersion using BCI-controlled mechanics. In Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology (pp. 1–6).

64 Wu, M.L., & Popescu, V. (2017). Efficient VR and AR navigation through multiperspective occlusion management. IEEE transactions on visualization and computer graphics, 24(12), 3069–3080.

65 Ylipulli, J., Luusua, A., & Ojala, T. (2017). On Creative Metaphors in Technology Design: Case" Magic". In Proceedings of the 8th International Conference on Communities and Technologies (pp. 280–289).

66 Yu, K., Eck, U., Pankratz, F., Lazarovici, M., Wilhelm, D., & Navab, N. (2022). Duplicated reality for co-located augmented reality collaboration. IEEE Transactions on Visualization and Computer Graphics, 28(5), 2190–2200.

67 Zhang, J., Dong, Z., Lindeman, R., & Piumsomboon, T. (2020). Spatial scale perception for design tasks in virtual reality. In Proceedings of the 2020 ACM Symposium on Spatial User Interaction (pp. 1–3).

Table B.1: List of reviewed publications in the literature review.

C Experimental Questionnaire

ID	Question	Low (left)	High (right)	Concept
I0	How much training does an average user need to perform this locomotion technique?	No training needed	Extensive training needed	Time until mastery
I1	Does an average user require external help to remember how to use this interaction technique during the interaction, for example, labeled buttons or tooltips?	User acts independently	User requires help	Full Internalization of Schema
I2	How much does an average user need to concentrate on the interaction to successfully use this locomotion technique?	Low level of concentration	High level of concentration	Cognitive Load
C0	How plausible is it to use this locomotion technique to move in reality?	Not plausible at all	Completely plausible	Experience of World
C1	How similar are the involved patterns of physical movements in VR to patterns of physical movement that can be used to travel in reality?	Completely different	Identical	Control Agency / Fidelity
C2	How does moving in VR using this technique feel compared to moving in reality?	Moving feels different	Moving feels the same	Experience of Self
E0	Is this form of locomotion physically less demanding or more demanding than walking in reality?	Lower physical demand	Higher physical demand	Physical Demand
E1	Does an average user have reduced capabilities for moving or enhanced capabilities (for example, moving faster than walking)?	Reduced capabilities	Enhanced capabilities	human-world relation
E2	Would an average user feel to be more powerful when they use this technology, or would he or she feel less powerful?	Feels less powerful	Feels more powerful	Subjectively Perceived Agency

Table C.1: The items used in the pilot questionnaire and their associated concept. For internalizability (I), congruence (C), and enhancement (E) three questionnaire items were formulated that measure aspects of the proposed dimensions.

D Description of Interaction Techniques

Point & Teleport The virtual environment shows a wide field without objects. The user holds one controller in his or her hand.

To move within the environment, the user can press the trigger on the controller to display a virtual ray that emerges from the controller. Next, the ray is pointed at a desired location using controller movements. As soon as the user releases the trigger, he or she is teleported to the indicated location.

Natural Walking The virtual environment shows a wide field without objects. The user holds two controllers in his or her hands, which are displayed within the virtual scene. The user's feet are not displayed.

The user walks in the virtual environment just like in the real world using leg movement.

Controller Flying In a seated position, the user holds a standard console controller (e.g., XBOX or Playstation) in his or her hands.

To fly within the virtual environment, the user presses the right trigger to increase the forward velocity. The direction of flight is controlled using head movements (tilt, yaw, roll), which are transferred to the camera's orientation. A neutral head position corresponds to flying straight forward. To slow down movement, the user presses the left trigger.

Finger Walking The user holds a multi-touch tablet in his or her hand. The virtual scene shows a virtual replication of the user's hands and the tablet.

To move within the environment, the user performs a walking motion with his or her index and middle finger on the tablet's surface by mimicking alternating steps with both fingers. To turn, the user presses the index and middle finger onto the tablet and rotates this finger configuration. As long as the user continues the walking pattern, the avatar moves forward at a constant speed of 4 km/h (walking speed).

Command-Line-Interface Using a menu button on a controller, the user opens a virtual menu that shows a text field input and a virtual keyboard.

To move within the virtual environment, a command following the system-specified syntax has to be typed in, such as: "teleport avatar swimming_pool". After confirming the text input using a virtual enter button, the user is teleported to the associated location.

Virtual Airplane The virtual scene shows an accurately modeled cockpit of a commercial airliner with buttons, switches, levers, gauges, and displays. The user's hands are tracked and visualized within the virtual environment. The user can grab the steering wheel with both hands and activate control elements using his or her fingertips.

To move within the virtual environment, the user has to control the virtual airplane just like a real-world airplane in manual flight mode (e.g., correct settings, correct start sequence, etc.). To change the flight direction, the user has to control the steering wheel accordingly.

Orb Teleportation The virtual environment shows a wide field without objects. The user holds two controllers in his or her hands, which are displayed within the virtual scene.

Using a menu button on a controller, four different spherical objects ('orbs') appear in front of the user. Each orb shows a rendered image of a different virtual location. The user can grab an orb by placing the controller within the orb and pressing the trigger button. When the user moves the orb towards his or her head and releases the trigger, the user is teleported to the location depicted on the surface of this orb.

Seven-League-Boots The virtual environment shows a wide field without objects. The user holds two controllers in his or her hands, which are displayed within the virtual scene. The user's feet are not displayed.

Users navigate within the virtual environment by moving their legs. The movement speed is increased by a factor of 3, so that one step in reality equals three steps in the virtual environment.

E Responses to the Pilot Questionnaire

Teleportation								
I0	I1	I2	C1	C2	C3	E0	E1	E2
3	3	2	1	1	2	2	7	6
2	2	1	1	1	1	1	5	6
3	2	3	1	1	1	1	7	6
3	2	2	1	1	2	3	5	4
2	1	1	3	2	1	1	6	5
2	6	3	1	1	1	1	6	5
2	1	3	1	1	1	2	6	5
4	4	2	1	2	2	1	7	6
2	2	3	1	1	2	2	6	4
2	1	2	2	4	1	1	4	6
2	2	3	2	3	1	1	5	4
3	2	3	1	1	1	1	7	7
1	1	2	2	1	1	1	4	5
1	1	2	3	1	1	1	4	5
2	3	3	1	1	1	1	5	5

Natural Walking								
I0	I1	I2	C1	C2	C3	E0	E1	E2
1	1	1	7	7	7	4	3	3
1	1	1	7	7	6	4	4	4
1	1	1	7	7	7	4	3	4
1	1	1	7	7	6	4	4	4
1	1	1	7	7	6	4	4	4
1	1	2	7	6	6	4	2	3
1	1	1	7	7	7	4	4	4
1	1	1	7	7	7	4	4	3
5	3	5	6	6	2	5	2	2
1	2	1	7	7	7	4	2	4
1	1	1	7	7	5	4	4	4
2	2	3	7	7	7	4	4	3
1	1	1	7	7	7	4	3	3
1	1	1	7	7	7	7	3	2
1	1	1	7	7	7	4	4	4

Controller Flying								
I0	I1	I2	C1	C2	C3	E0	E1	E2
3	3	3	3	2	2	2	6	5
5	3	4	2	1	3	2	6	5
4	2	2	1	1	1	1	7	7
4	3	3	2	1	2	2	5	5
4	2	4	1	1	1	2	6	6
6	7	6	2	1	1	5	6	4
4	6	6	2	2	2	2	6	6
4	2	5	2	2	2	4	3	3
6	5	6	2	4	2	5	2	4
6	7	6	5	6	5	4	4	4
3	2	4	5	3	3	1	4	4
5	4	6	5	1	1	1	3	5
4	2	5	5	3	1	1	7	6
3	3	4	3	3	3	4	6	7
3	2	5	1	1	1	2	6	6

Finger Walking								
I0	I1	I2	C1	C2	C3	E0	E1	E2
3	4	3	2	3	2	3	4	3
7	5	6	1	1	1	2	2	1
5	4	4	1	2	1	2	4	2
2	2	3	1	2	2	2	3	3
4	3	4	1	1	1	2	2	3
4	2	5	1	5	2	3	3	3
3	4	4	1	1	1	2	4	3
2	4	4	5	4	2	2	3	3
6	5	5	2	2	2	5	2	3
6	7	6	1	1	1	6	2	1
6	5	6	5	5	2	2	2	2
7	6	6	3	1	1	1	1	1
3	4	5	2	1	1	1	2	2
5	5	6	1	1	1	5	2	2
3	6	5	1	2	1	2	2	1

Command-Line Interface								
I0	I1	I2	C1	C2	C3	E0	E1	E2
5	4	5	1	1	1	1	6	4
7	7	7	2	1	1	1	5	2
5	6	7	1	1	1	2	1	2
6	7	6	1	1	1	3	5	4
5	3	4	1	1	1	1	5	3
6	7	6	1	1	1	1	2	2
6	7	6	2	2	2	1	4	5
5	4	4	1	1	1	1	7	7
7	7	7	2	2	2	1	2	2
7	7	6	1	1	1	4	2	5
7	6	7	4	1	1	2	5	2
7	7	7	2	1	1	1	7	1
2	6	6	3	3	1	1	2	3
6	6	6	1	1	1	5	2	2
6	7	7	1	1	1	2	2	3

Realistic Airplane								
I0	I1	I2	C1	C2	C3	E0	E1	E2
7	5	6	7	7	6	2	6	4
7	7	6	7	3	3	1	3	4
7	7	7	5	7	7	4	5	6
7	6	6	4	6	5	2	5	4
7	7	7	1	3	3	2	5	6
5	7	6	6	7	5	2	6	7
7	7	7	7	6	6	4	4	4
7	7	7	7	7	5	3	5	7
7	7	7	7	7	7	2	1	1
7	7	6	6	6	6	4	4	4
7	7	7	6	6	5	7	2	2
7	6	6	7	1	3	1	3	5
6	5	6	7	6	6	2	6	6
7	7	7	7	7	7	7	5	7
7	7	7	7	7	6	2	6	4

Orb Teleportation								
I0	I1	I2	C1	C2	C3	E0	E1	E2
2	2	3	1	1	1	3	6	5
4	3	3	1	1	1	1	4	4
4	4	4	1	1	1	2	7	7
2	2	2	1	1	1	2	5	5
2	1	1	1	1	1	1	5	4
4	3	3	1	1	1	3	5	5
2	2	2	1	1	1	2	6	6
3	3	2	1	1	1	1	7	7
3	3	4	1	1	2	2	5	5
6	6	6	1	1	1	2	2	5
3	4	1	5	2	1	1	6	5
4	2	4	1	1	1	1	2	4
2	2	1	3	1	1	1	4	4
3	5	3	1	1	1	6	3	3
3	5	5	1	1	1	2	5	4

7-League-Boots								
I0	I1	I2	C1	C2	C3	E0	E1	E2
2	1	1	3	6	3	4	5	6
2	1	2	5	6	6	3	5	5
2	1	1	6	7	6	3	6	5
3	2	2	3	6	5	3	5	5
2	1	2	2	2	5	2	5	5
2	1	2	4	7	5	3	5	4
2	1	2	6	6	6	3	7	7
1	1	2	6	7	4	2	7	7
6	2	5	5	5	2	5	2	2
5	4	5	5	5	5	6	6	6
3	2	4	6	6	4	3	5	4
3	2	6	3	1	1	1	4	4
2	1	2	5	6	5	2	6	5
1	1	1	6	7	7	2	6	6
4	1	4	4	6	5	4	5	5

Figure 1: Raw answers to the pilot questionnaire.

Answers are encoded into values ranging from 0 to 7, with 0 corresponding to the 'low' answer and 7 corresponding to the 'high' answer.

		TP	NW	CF	FW	CLI	AP	Orb	7LB
Internalizabil.	*max*	94	100	72	78	50	22	94	100
	$Q_{.75}$	89	100	61	56	28	11	81	92
	median	78	100	56	44	11	0	61	89
	$Q_{.25}$	72	97	31	25	6	0	58	67
	min	61	44	11	11	0	0	17	39
Congruence	*max*	22	100	72	50	28	100	11	94
	$Q_{.75}$	11	100	31	19	8	92	0	78
	median	6	100	17	6	0	89	0	72
	$Q_{.25}$	0	94	8	0	0	72	0	56
	min	0	61	0	0	0	22	0	11
Enhancement	*max*	100	50	100	42	100	92	100	100
	$Q_{.75}$	83	50	83	33	54	79	75	79
	median	75	42	75	25	42	67	67	67
	$Q_{.25}$	63	33	50	13	21	50	50	63
	min	58	17	33	0	8	0	33	17

Table E.1: Descriptive statistics of the calculated ratings for internalizability, congruence, and enhancement (min, 25%-quartile, median, 75%-quartile, max).

F Questionnaires

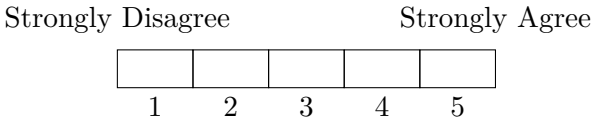
System Usability Scale (SUS)

Measurement

A scale with 5 segments. At the end of each scale, the semantic differentials 'Strongly Disagree' - 'Strongly Agree' are presented. The segments of the scale are labeled with 1 to 5. The SUS score is calculated by adding the values of items 1, 3, 5, 7, and 9 (each minus 1), added to the sum of the inverted values of items 2, 4, 6, 8, and 10. The obtained value is multiplied by 2.5 to obtain the overall value of SUS between 0 and 100 [Bro+96].

Instructions

Please rate each statement according to your level of agreement using the scale below:



Questions

Q1. I think that I would like to use this system frequently.
Q2. I found the system unnecessarily complex.
Q3. I thought the system was easy to use.
Q4. I think that I would need the support of a technical person to be able to use this system.
Q5. I found the various functions in this system were well integrated.
Q6. I thought there was too much inconsistency in this system.
Q7. I would imagine that most people would learn to use this system very quickly.
Q8. I found the system very cumbersome to use.
Q9. I felt very confident using the system.
Q10. I needed to learn a lot of things before I could get going with this system.

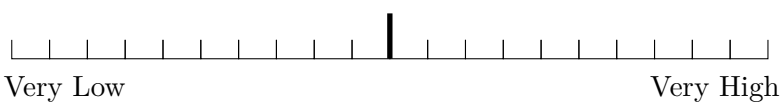
NASA Task Load Index (NASA-TLX)

Measurement

A scale with 20 segments capturing the subjective experience of six different aspects. At the end of each scale, semantic differentials are presented. In the case of Performance, 'Perfect' - 'Failure', in all other cases, 'Very Low' - 'Very High'. It is possible to calculate an *overall workload value*, or to compare the different subscales individually [Har06].

Instructions

Please rate each of the following subscales according to your experience using the scale below:



Questions

Q1. (*Mental Demand*)) How much mental and perceptual activity was required? Was the task easy or demanding, simple or complex?

Q2. (*Physical Demand*) How much physical activity was required? Was the task easy or demanding, slack or strenuous?

Q3. (*Temporal Demand*) How much time pressure did you feel due to the pace at which the tasks or task elements occurred? Was the pace slow or rapid?

Q4. (*Performance*) How successful were you in performing the task? How satisfied were you with your performance?

Q5. (*Effort*) How hard did you have to work (mentally and physically) to accomplish your level of performance?

Q6. (*Frustration*) How irritated, stressed, and annoyed versus content, relaxed, and complacent did you feel during the task?

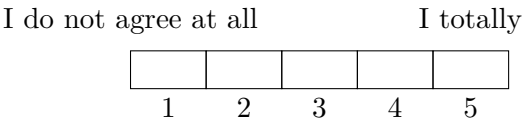
Spatial Presence Experience Scale (SPES)

Measurement

A scale with 5 segments. At the end of each scale, the semantic differentials 'Strongly Disagree' - 'Strongly Agree' are presented. The segments of the scale are labeled with 1 to 5. Depending on the focus of the research, the overall score can be compared or the score of the subscales *self-location* and *possible actions* [Har+15].

Instructions

Please rate each of the following statements according to your experience (1 = I do not agree at all, 5 = I totally agree)



Questions

SL1. I felt like I was actually there in the environment of the presentation.

SL2. It seemed as though I actually took part in the action of the presentation.

SL3. It was as though my true location had shifted into the environment in the presentation.

SL4. I felt as though I was physically present in the environment of the presentation.

PA1. The objects in the presentation gave me the feeling that I could do things with them.

PA2. I had the impression that I could be active in the environment of the presentation.

PA3. I felt like I could move around among the objects in the presentation.

PA4. It seemed to me that I could do whatever I wanted in the environment of the presentation.

DECLARATION

Eidesstattliche Versicherung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Sofern im Zuge der Erstellung der vorliegenden Dissertationsschrift generative Künstliche Intelligenz (gKI) basierte elektronische Hilfsmittel verwendet wurden, versichere ich, dass meine eigene Leistung im Vordergrund stand und dass eine vollständige Dokumentation aller verwendeten Hilfsmittel gemäßder Guten wissenschaftlichen Praxis vorliegt. Ich trage die Verantwortung für eventuell durch die gKI generierte fehlerhafte oder verzerrte Inhalte, fehlerhafte Referenzen, Verstöße gegen das Datenschutz- und Urheberrecht oder Plagiate.

Declaration on oath

Hereby, I declare under oath that I have written the present dissertation myself and have not used any sources or aids other than those specified. Insofar as generative artificial intelligence (gAI)-based electronic tools were used in the course of preparing the present dissertation, I affirm that my own work remained the primary focus and that complete documentation of all tools used is available in accordance with good scientific practice. I accept responsibility for any faulty or distorted content, erroneous references, violations of data protection and copyright law, or plagiarism generated by the gAI.

Hamburg, den Unterschrift