University of Bremen

Unraveling the Fidelity of Virtual Reality Interactions

Effects of Realism in Object Manipulation and Embodiment

Michael D. Bonfert

First Reviewer: Prof. Dr. Rainer Malaka
Second Reviewer: TT-Prof. Dr. Tiare Feuchtner

Digital Media Lab
Faculty 3: Mathematics & Computer Science
University of Bremen

Submitted on 27th February, 2024
Defended on 12th April, 2024

Doctoral dissertation submitted in partial fulfillment
of the requirements for the degree of Doctor of Engineering (Dr.-Ing.)

Second edition, Tuesday 16th April, 2024

University of Bremen

Statement of Authorship

I hereby declare that I am the legitimate author of this dissertation and that it is my original work. All direct or indirect sources used are acknowledged as references. I confirm that I have written this thesis without illegitimate help and that I have not used any sources and aids other than permitted in the *ACM Policy on Authorship*. No portion of this work has been submitted in support of an application for another degree or qualification of this or any other university or institution of higher education.

Candidate	Michael D. Bonfert
Student ID	3032695
Faculty	Faculty 3: Mathematics & Computer Science

Signature	______________________
City, date	Bremen, Tuesday 16th April, 2024

Preface

The more I worked with virtual and mixed reality, the more clearly I realized that interactions with digital content will likely become spatially embedded into our natural environments. I expect a shift from 2D desktops and smartphones to 3D interfaces projected into our surroundings as the next central HCI paradigm—even if we still have a long way to go. I am glad I got the opportunity to glimpse into this possible future and help shape these interfaces by better understanding how people use them. The journey to this dissertation was incredibly rewarding. I tremendously enjoyed this chapter of my life, learned so much, and have grown with the challenges. I want to express my gratitude to the many people who supported me on my way and contributed to this work.

First, I want to thank my doctoral supervisor, Rainer Malaka, for always having my back. You offered me the opportunity and gave me the security to pursue my doctorate. As the director of the Digital Media Lab, you provided me with a friendly and supportive work environment. I appreciate that you let me follow my curiosities with incredible freedom in my research. I further thank Tanja Döring for guiding me through the obstacles and uncertainties I encountered in my studies and the academic world. Your supervision gave me direction, stability, and confidence. Many thanks to Tiare Feuchtner for reviewing my thesis and Udo Frese, René Weller, Rachel Ringe, and Shiyao Zhang for joining the examination committee.

I am thankful to all my collaborators and co-authors for the fruitful discussions and efforts in jointly shaping our research. You were at the core of my daily work, and I very much enjoyed doing research with you. I had the privilege to work with dedicated and talented students, including Maiko Hübinger, Malte Borgwardt, Jonas Boueke, Florian Saade, and Stella Lemke. Thanks for your trust and exceptional commitment. Special thanks go to all my fantastic (former) colleagues at the Digital Media Lab. I appreciate our discussions over coffee, lunch breaks, climbing together, traveling together, the awesome retreats, and the relaxed atmosphere in our lab. I'm especially grateful to Anke, Thomas, Nima, Johanna, Bastian, Susanne, and Caro, with whom I shared not only great projects and inspiring conversations but also friendships. Many thanks to Svenja and Evgenia for their tireless and patient assistance, Philipp and Irmgard for their competent technical support, and Robert and Gerald for their valuable advice and guard rails along the way. What a team!

It was a pleasure being introduced to the academic HCI community, where I met countless wonderful and warm-hearted peers. I am thankful to you all for exchanging ideas, sharing helpful feedback, volunteering to organize conferences, and creating an atmosphere of appreciation at scientific events. I value the time and consideration of the senior researchers who shared their experiences with me. I further thank the Klaus Tschira Stiftung for generously funding my scholarship and providing an extraordinary community of scholars. You allowed me to focus on the research close to my heart.

Finally, I want to express my deepest gratitude to my family and friends. I thank my close and far companions for their emotional support, interest, and soothing distractions. In particular, I am thankful to Maike and Maren, who were always there for me and supported me unconditionally with all their hearts and brilliance. You enabled me to accomplish this. I want to thank Wojtek for encouraging me and being there for me. And many thanks to my wonderful friends Jenny, Moritz, Liz, Jie, Mette, Anna, Lotti, Tina, Andy, Anne, Kathrin, Max, my Mensa peers, the Improv-AMSeln, and the MuKlers. You made it such an enjoyable chapter of my life and helped me find balance. I am most lucky and grateful to have Jan-Georg by my side. His compassion and love are an endless source of courage and joy. Thanks for asking, listening, and understanding, for believing in me, for being close and giving me space, for joking and laughing together, and for being my haven of peace. Also, I would like to thank my dear brother Stefan, his wife Christina, and my grandparents for their caring support and encouragement. Above all, I am immensely grateful to my parents, Brigitte and Jürgen, for their confidence and support in reaching my goals. I could count on you at every step that brought me here. You were there for me with patience, an open ear, advice, and endless love.

Abstract

Immersive experiences created with multisensory and interactive virtual reality (VR) technology can feel incredibly realistic. Many factors of the system and the user determine the fidelity of VR interactions, that is, how closely an interaction resembles the reference interaction to be simulated. The human-computer interaction (HCI) research presented in this dissertation is a step towards unraveling and better understanding the different aspects of interaction fidelity in VR that cover user input, computer simulation, sensory feedback, and the user's experience.

Combining theoretical, methodological, and empirical research, I present and discuss nine papers focusing on object manipulation and embodiment in the context of interaction fidelity. The presented *Interaction Fidelity Model* enables a systematic assessment of eight fidelity components with precise definitions and thorough discussion of their implications. Integrating the HCI loop and previous theories, the model's human-centered perspective can be universally applied to all interactions with any VR system. Furthermore, the specialized *Haptic Fidelity Framework* allows a detailed assessment of 14 factors determining the fidelity of haptic interfaces.

On this theoretical basis, a series of user studies are analyzed, covering findings on haptic feedback and perception for object manipulation, body visibility in VR sports, visual realism of virtual humans, embedding questionnaires into virtual environments for user research, and VR meetings of an academic team. The quantitative and qualitative analyses demonstrate significant effects of varying interaction fidelity on user performance, task load, sense of body ownership, perceived realism, usability, and user experience. The iterative user research contributes hardware and software prototypes, such as an in-VR questionnaire tool for participants' self-reports, a grip-based interaction technique for force-feedback gloves, a handheld controller with adaptive trigger resistance, and experimental environments built with Unity.

This thesis provides guidelines, illustrative examples, research artifacts, and educational material for designing realistic VR applications, improving how we comprehend and reflect on interaction fidelity in VR research and development. An extensive research agenda highlights promising directions for future investigations of the complex construct of interaction fidelity.

Zusammenfassung

Immersive Erlebnisse durch multisensorische und interaktive Virtual Reality (VR) können sich erstaunlich realistisch anfühlen. Verschiedene System- und Nutzer-Faktoren bestimmen die Fidelity (Originaltreue) von VR-Interaktionen, also wie sehr eine Interaktion der zu simulierenden Referenzinteraktion ähnelt. Die in dieser Dissertation vorgestellte Forschung im Bereich der Mensch-Computer-Interaktion (HCI) ist ein Schritt zum tieferen Verständnis der Aspekte der Interaktionstreue in VR, die Input, Computersimulation, sensorisches Feedback und Benutzererfahrung umfassen.

Mit einer Kombination aus theoretischer, methodischer und empirischer Forschung präsentiere ich neun Arbeiten mit Fokus auf Objektmanipulation und Embodiment im Kontext der Interaktionstreue. Das vorgestellte *Interaction Fidelity Model* ermöglicht eine systematische Evaluierung von acht Fidelity-Komponenten mit präzisen Definitionen und umfassender Diskussion. Durch die Integration der HCI-Schleife und früherer Theorien kann die menschenzentrierte Perspektive des Modells universell auf alle Interaktionen mit beliebigen VR-Systemen angewendet werden. Das spezialisierte *Haptic Fidelity Framework* ermöglicht eine detaillierte Bewertung von 14 Faktoren, die die Fidelity von haptischen Systemen beeinflussen.

Auf dieser theoretischen Grundlage werden Nutzerstudien analysiert, die Erkenntnisse umfassen über haptisches Feedback bei Objektmanipulationen, Visualisierung des Körpers im VR-Sport, visuellen Realismus virtueller Menschen, Einbettung von Fragebögen in VR und Team-Meetings in VR. Die quantitativen und qualitativen Untersuchungen zeigen Effekt von Fidelity auf Nutzer-Performance, Arbeitsbelastung, Body Ownership, empfundenen Realismus, Usability und User Experience. Aus der iterativen Nutzerforschung gehen Hardware- und Software-Prototypen hervor, darunter ein VR-Fragebogen-Tool, eine Interaktionstechnik für Force-Feedback-Handschuhe, ein Controller mit adaptivem Trigger-Widerstand und Experimentumgebungen in Unity.

Diese Doktorarbeit enthält Leitlinien, anschauliche Beispiele, Forschungsartefakte und Lehrmaterial für die Gestaltung realistischer VR-Anwendungen und fördert unser Wissen über Interaktionstreue in der VR-Forschung und -Entwicklung. Eine umfangreiche Forschungsagenda zeigt vielversprechende Richtungen für zukünftige Untersuchungen des komplexen Konstrukts der Fidelity auf.

Contents

List of Figures

List of Abbreviations

3DUI 3-dimensional user interface

ACM Association for Computing Machinery

AR Augmented reality

CHI Conference on Human Factors in Computing Systems

CRediT Contributor Roles Taxonomy

CUI Conversational user interfaces

DOF Degrees of freedom

DOI Digital Object Identifier

HCI Human-computer interaction

HMD Head-mounted display

IEEE Institute of Electrical and Electronics Engineers

inVRQ in-VR questionnaire

MR Mixed reality

outVRQ out-VR questionnaire

RQ Research question

TLX Task Load Index

UI User interface

UX User experience

VE Virtual environment

VR Virtual reality

XR xReality, an umbrella term for AR, MR, and VR

Part I

Thesis

1
Introduction

The rapid evolution of virtual reality (VR) technology enables simulating virtual worlds that feel increasingly realistic. Today, immersive VR experiences can already be profoundly captivating by surrounding the user with the multimodal sensations of an interactive virtual environment (VE). This trend is set to continue and revolutionize our relationship with virtuality. Augmented reality (AR), mixed reality (MR), and VR—jointly encompassed under the umbrella term xReality (XR) (Rauschnabel et al., 2022)—have the potential to fundamentally transform how we access digital information and integrate virtual content into our lives. A human-centered approach based on the methodological and ethical principles of human-computer interaction (HCI) ensures that XR systems empower users and positively impact society and individuals.

Virtual worlds created with VR technology can be described and evaluated with various metrics and goals. They can be immersive, enjoyable, colorful, comfortable, demanding, sustainable, creative, effective, frightening, or purposeful, among many other attributes. One essential characteristic is their realism. For many VR applications, this is the decisive quality metric. The true-to-life resemblance is fundamental in many use cases: skill training, e.g., surgery (Chheang et al., 2019); learning abilities, e.g., in sports (Schulz et al., 2019) or music (Serafin et al., 2017); vocational education, e.g., public speaking (Poeschl, 2017); entertainment, e.g., traveling the world (Sarkady et al., 2021); therapy, e.g., fear of heights (Freeman et al., 2018); or use cases that are expensive or impossible without VR, e.g., visiting Mars (Holt, 2023). In these scenarios, the success of the simulation depends on how closely they can reproduce reality. Even in fictional scenarios, some system aspects might need to be grounded in reality, such as Euclidean geometry, spatial audio, swarm behavior, gravity, or the color space humans perceive.

Simulated worlds share similarities with reality in some aspects but differ in others. The level of realism is determined by various factors, including system design, device specifications, user characteristics, interaction techniques, application goals, and other parameters or circumstances. It is also important in interaction design to consider the effects of varying realism on user perception, experience, and behavior. All these factors make realism a complex and multilayered concept.

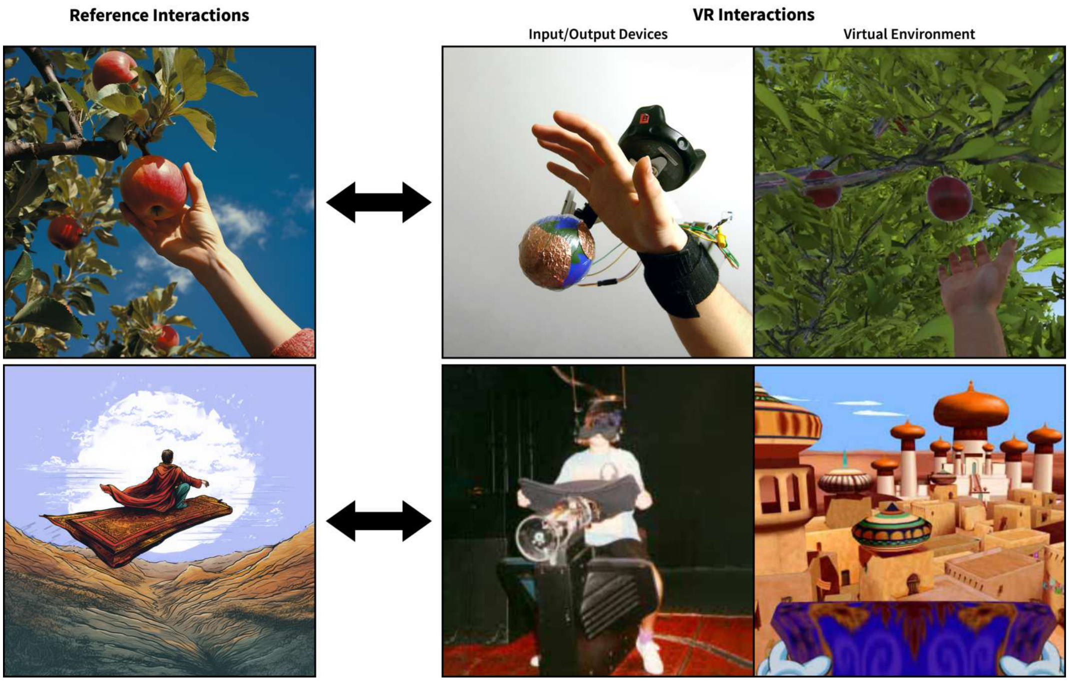


Figure 1.1: Two examples of reference interactions that are virtually replicated in VR. *(top)* Picking a real apple from a tree is reproduced with an implementation using a haptic device by de Tinguy et al. (2020), dynamically moving the sphere proxy into the user's hand when grasping. *(bottom)* The fictional *magic carpet* from Middle Eastern literature is reproduced with an early implementation by Pausch et al. (1996). *Figure from **P1** (IntFi Model). Copyright of the top-right images by de Tinguy et al. (2019) and the bottom-right images by Pausch et al. (1996) (modified). The images for the reference interactions have been generated with Midjourney.*

The concept of realism is inherently limited by describing a simulation's correspondence to reality. More flexibly, *fidelity* describes how accurately a simulation reproduces any reference (Gerathewohl, 1969; Merriam-Webster, 2024) beyond the imitation of the real world. Thus, realism can be seen as one specific form of fidelity: fidelity to reality. Other reference frames may include fictional elements, supernatural abilities, dreams, or ideas outside the bounds of reality. For instance, although it is unrealistic, flying like a superhero can be virtually realized with high fidelity.

In this thesis, I address the interactions between users and VR systems, similar to how people constantly interact with the real world. Interaction fidelity describes the exactness of correspondence between these two. The actions taken by users and the output generated by a computer depend on each other and together form the interaction. It is important to consider them together (Hornbæk and Oulasvirta, 2017). To describe, analyze, or assess interaction fidelity, we compare the VR interactions to the reference to be simulated, as illustrated in Figure 1.1 with a real-life and a fictional example. However, interaction fidelity is a complex and manifold concept because countless human and technological factors determine it. To purposefully design virtual experiences that are convincingly realistic, it is essential to untangle the different aspects that impact the overall realism.

1.1 | Research Questions

This doctoral dissertation contributes to a deeper understanding of VR interactions with a focus on aspects of fidelity. It presents a combination of theoretical works, methodological research, the development of research prototypes, and their experimental evaluations with user studies to unravel the complexity of interaction fidelity in VR. The included user studies cover topics around haptics, object manipulation, embodiment, and the use of questionnaires in VR user studies. Applying the proposed Interaction Fidelity Model and Haptic Fidelity Framework to this wide thematic range illustrates how universal and valuable the concept of fidelity is in interaction design. The following research questions guide this work.

RQ1: How can we structure and describe the aspects determining the fidelity of interactions in virtual reality?

RQ2: How can conventional user study questionnaires be transferred to VR with high fidelity?

RQ3: What are the effects of varying interaction fidelity for manipulating virtual objects on perception and performance?

RQ4: What are the effects of varying interaction fidelity on how the embodiment of others and oneself is perceived?

I summarize and discuss the publications in this dissertation in the context of these overarching questions. In my work, I followed an iterative and human-centric approach. My research is based on various quantitative and qualitative empirical research methods, including expert interviews, literature reviews, psychophysical experiments, and user evaluations with objective and subjective measures, including behavioral data and self-reports. Integrating theory, prototype development, and empirical evaluation ensures a comprehensive and valid investigation of VR interactions.

1.2 | Included Publications

This *dissertation by publication* comprises eight peer-reviewed publications: four full papers presented at international conferences (ACM CHI, IEEE VR, ACM CUI), three full articles in journals (Frontiers in VR), and one late-breaking work (ACM CHI). These publications are included in their original format in Part II as published. Additionally, this dissertation includes a full paper on the Interaction Fidelity Model, **P1** (IntFi Model), that was under review when this thesis was submitted. The paper was previously submitted to ACM CHI, where it received positive feedback and suggestions for improvement that have since been implemented. The paper is included in Part II as an arXiv preprint.

The following chapters of Part I will briefly introduce each paper accompanied by a fidelity analysis based on the theoretical works in response to the research questions. The works are thematically organized into four chapters: Chapter 2 contains two theoretical papers on the fidelity of VR interactions (**P1**, **P2**), Chapter 3 covers methodological work on questionnaires in VR research (**P3**), Chapter 4 includes two empirical studies on object manipulation in VR (**P4**, **P5**), and Chapter 5 presents four empirical papers on embodiment in VR (**P6–P9**). While these summaries cannot substitute reading the publications, they give an overview of the research to facilitate the overarching discussion in the context of the dissertation's thematic focus. For this, I address the research questions by discussing the papers regarding fidelity at the end of each chapter. In the subsequent discussion in Chapter 6, I contextualize the findings and give an outlook, followed by a conclusion in Chapter 7. The original publications in Part II provide the complete study designs, detailed results, statistical reports, and in-depth discussions in the context of related work beyond the concept of fidelity. The following publications form the foundation of this thesis.

P1 Michael Bonfert, Thomas Muender, Ryan P. McMahan, Frank Steinicke, Doug Bowman, Rainer Malaka, and Tanja Döring. 2024. **The Interaction Fidelity Model: A Taxonomy to Distinguish the Aspects of Fidelity in Virtual Reality.** Preprint on arXiv. arXiv:2402.16665v1 [cs.HC]

P2 Thomas Muender, Michael Bonfert, Anke Reinschluessel, Rainer Malaka, and Tanja Döring. 2022. **Haptic Fidelity Framework: Defining the Factors of Realistic Haptic Feedback for Virtual Reality.** In Proceedings of the 2022 Conference on Human Factors in Computing Systems (CHI '22), Association for Computing Machinery. DOI: 10.1145/3491102.3501953

P3 Dmitry Alexandrovsky, Susanne Putze, Michael Bonfert, Sebastian Höffner, Pitt Michelmann, Dirk Wenig, Rainer Malaka, and Jan David Smeddinck. 2020. **Examining Design Choices of Questionnaires in VR User Studies.** In Proceedings of the 2020 Conference on Human Factors in Computing Systems (CHI '20), Association for Computing Machinery. DOI: 10.1145/3313831.3376260

P4 Michael Bonfert, Maiko Hübinger, and Rainer Malaka. 2023. **Challenges of Controlling the Rotation of Virtual Objects with Variable Grip Using Force-Feedback Gloves.** Frontiers in Virtual Reality 4. DOI: 10.3389/frvir.2023.1190426

P5 Carolin Stellmacher, Michael Bonfert, Ernst Kruijff, and Johannes Schöning. 2022. **Triggermuscle: Exploring Weight Perception for Virtual Reality Through Adaptive Trigger Resistance in a Haptic VR Controller.** Frontiers in Virtual Reality 2. DOI: 10.3389/frvir.2021.754511

P6 Michael Bonfert, Stella Lemke, Robert Porzel, and Rainer Malaka. 2022. **Kicking in Virtual Reality: The Influence of Foot Visibility on the Shooting Experience and Accuracy.** In Proceedings of the 2022 Conference on Virtual Reality and 3D User Interfaces (IEEE VR '22), IEEE Computer Society. DOI: 10.1109/VR51125.2022.00092

P7 Malte Borgwardt, Jonas Boueke, María Fernanda Sanabria, Michael Bonfert, and Robert Porzel. 2023. **VRisbee: How Hand Visibility Impacts Throwing Accuracy and Experience in Virtual Reality.** In Extended Abstracts of the 2023 Conference on Human Factors in Computing Systems (CHI EA '23), Association for Computing Machinery. DOI: 10.1145/3544549.3585868

P8 Michael Bonfert, Nima Zargham, Florian Saade, Robert Porzel, and Rainer Malaka. 2021. **An Evaluation of Visual Embodiment for Voice Assistants on Smart Displays.** In Proceedings of the 2021 Conference on Conversational User Interfaces (CUI '21), Association for Computing Machinery. DOI: 10.1145/3469595.3469611

P9 Michael Bonfert, Anke V. Reinschluessel, Susanne Putze, Yenchin Lai, Dmitry Alexandrovsky, Rainer Malaka, and Tanja Döring. 2023. **"Seeing the faces is so important"—Experiences from online team meetings on commercial virtual reality platforms.** Frontiers in Virtual Reality 3. DOI: 10.3389/frvir.2022.945791

All research has been realized collaboratively in teams. Therefore, I will use "we" to report on it. Only in exceptions will I use "I" when referring to my work on this thesis. At the beginning of the paper summaries in Part I and before the original papers in Part II, I will state my individual contributions to the included works, using the Contributor Roles Taxonomy *CRediT*[1] to distinguish them from my co-authors' part. Chapter 9 provides a complete list of my publications, including works outside the scope of this dissertation. These references link to the original online resources via the Digital Object Identifier (DOI).

[1] https://credit.niso.org/contributor-roles-defined

2

Fidelity of VR Interactions

One way to conceptualize VR systems is by evaluating their resemblance to the reference they aim to simulate. Every VR simulation is designed to recreate some reference, be it a real-life situation, a training scenario, a fictional world, or a designer's imagination. If it replicates (parts of) the real world, we can assess its level of *realism*. More broadly speaking, we can use the concept of *fidelity* to describe how accurately a reference is reproduced (Merriam-Webster, 2024), be it reality or other reference frames. The term originates from the Latin word *fidēlis* for "faithful." Thus, realism is a specific form of fidelity. Therefore, a VR application can have low realism yet high fidelity to a reference frame other than the real world. For instance, the swords in the game *Beat Saber*[1] have high fidelity to lightsabers from *Star Wars* but are unrealistic. Raser suggested in 1969 to compare simulations more generally to various reference systems than solely to the real world, and we have adopted this notion for universal applicability in VR research.

Unfortunately, scientific literature often uses the terms fidelity and realism universally without specifying which aspect is being referred to. For example, the term *interaction fidelity* has been used to refer to visual render quality (Mania et al., 2006), camera views and gravity (Bhargava et al., 2018), or dialogue capabilities (Carnell et al., 2022). Some publications refer only to the user's system input with it (Bowman et al., 2012; McMahan et al., 2016), which neglects half the two-way interaction between the user and the system. Even when a specific aspect of fidelity is mentioned, a clear definition is often missing (Rogers et al., 2022). As a result, the definitions and terms used in VR literature have been inconsistent and contradictory.

On the other hand, the community's research efforts have resulted in valuable definitions, models, and frameworks in prior research (Al-Jundi and Tanbour, 2022; Alexander et al., 2005; Beckhaus and Lindeman, 2011; Jacob et al., 2008; McMahan et al., 2016; Stoffregen et al., 2003). While these works cover a variety of factors and concepts, they mainly focus on a few selected aspects of fidelity, use different terms, and cannot provide a comprehensive understanding. In their systematic review of realism and fidelity

[1] https://beatsaber.com

for digital games, Rogers et al. (2022) found a "substantial potential for confusion given the overlapping and contradictory use of realism types." The authors outlined the vast range of terms used to describe aspects of realism and fidelity.

While the concept of fidelity can be useful and allow insightful analyses, it can also be misleading if oversimplified. When describing or analyzing interaction fidelity in VR, it does not do the complexity of the construct justice to only gauge the screen resolution and score of a presence questionnaire. Although such approaches can be found frequently in the literature with the attempt to generalize study findings in the context of more or less fidelity, they do not allow for a comprehensive assessment of the multifaceted construct. Without specifying the investigated or experimentally manipulated fidelity component, oversimplified conclusions might be misleading and not advance our understanding of this complex construct as a more nuanced consideration would.

All this makes it difficult to establish links between individual discoveries, generalize the results, and synthesize fundamental principles. There is a need for a unified model that consolidates existing findings on different aspects of fidelity into one comprehensive and consistent taxonomy. This chapter addresses the first research question based on the publications **P1** (IntFi Model) and **P2** (Haptic Fidelity).

RQ1: How can we structure and describe the aspects determining the fidelity of interactions in virtual reality?

2.1 | The Interaction Fidelity Model

This section is based on publication **P1 (IntFi Model)**:

> Michael Bonfert, Thomas Muender, Ryan P. McMahan, Frank Steinicke, Doug Bowman, Rainer Malaka, and Tanja Döring. 2024. **The Interaction Fidelity Model: A Taxonomy to Distinguish the Aspects of Fidelity in Virtual Reality.** Preprint on arXiv.

My contribution to this work: I contributed the research idea and the majority of the theoretical considerations (*conceptualization*), the literature review, and most of the writing of the manuscript (*writing – draft, review, & editing*). I designed and conducted the validation method (*data curation* and *methodology*), conducted the expert interviews (*investigation*), and analyzed the qualitative data (*formal analysis*). I created most of the figures and the slide deck, and contributed to the other figures and the poster (*visualization*). I coordinated the project (*project administration*).

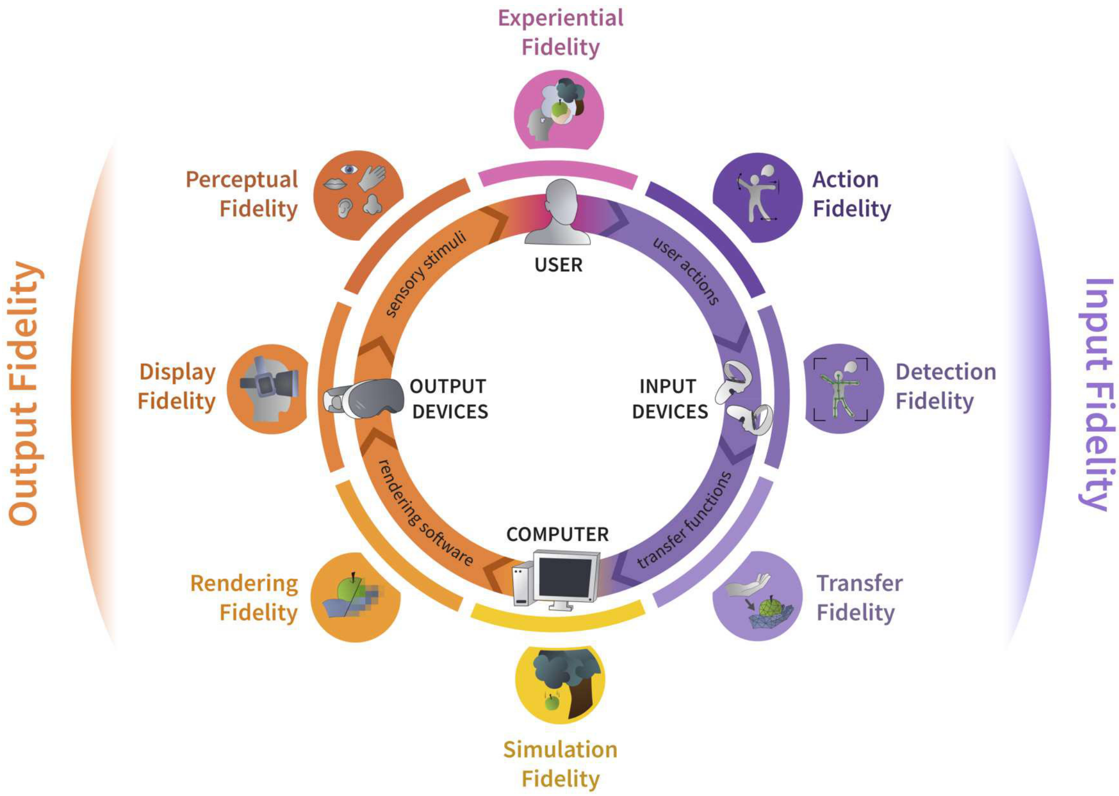


Figure 2.1: The Interaction Fidelity Model differentiates between eight components that determine the fidelity of any VR interaction.

To facilitate a complete and clear understanding of interaction fidelity, we created the Interaction Fidelity Model (IntFi Model). This conceptual model distinguishes the various aspects of fidelity inherent in all VR interactions. It can serve as an analytical scalpel to dissect the construct. The model encompasses the entire process of a user interacting with a VR system, from user input and system processing to the system output experienced by the user.

Instead of assessing the contribution of each device or system component to the fidelity of the interaction, we propose distinguishing between the stages of the interaction to evaluate systematically how true it is to the original. The IntFi Model is based on the HCI loop (Norman and Draper, 1986), which describes the stages of the user, input devices, computer, and output devices, illustrated with the inner circle in Figure 2.1. Our model expands on the Framework for Interaction Fidelity Analysis by McMahan et al. (2016), originally covering only the three components of input fidelity. Following the loop structure, the model assigns one aspect of fidelity (for example, display fidelity) to one stage of the loop (in this example, output devices), as illustrated in Figure 2.1. The

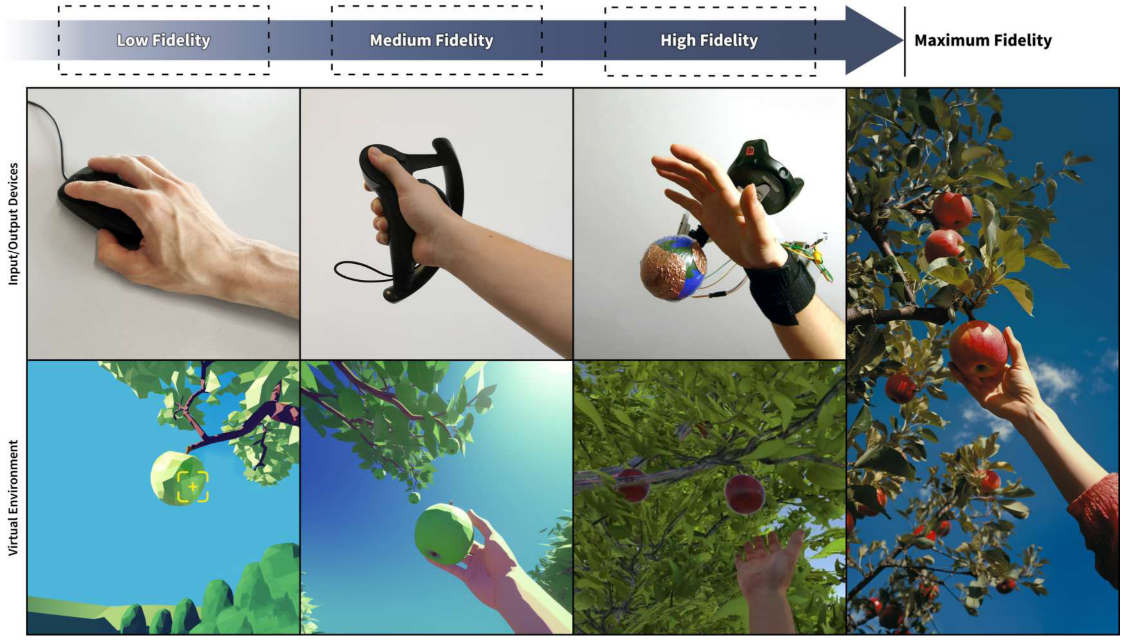


Figure 2.2: The fidelity spectrum with approximate ranges from low, medium, and high to maximum fidelity with an example use case: implementations with different fidelity levels of somebody picking an apple from a tree. The reference interaction from the real world on the right side is defined as maximum fidelity. *Copyright of the "High Fidelity" images by de Tinguy et al. (2020). The other images have been photographed or generated with Midjourney by the author.*

loop offers simplicity while integrating all fidelity aspects for any conceivable interaction. To make a meaningful and unambiguous assessment of how faithful VR interactions are to a reference interaction, the reference must be clearly defined.

We can describe the level of fidelity on a spectrum ranging from low, medium, and high to maximum fidelity as illustrated in Figure 2.2. Maximum fidelity represents a perfect correspondence to the original, even if it might be technologically impossible to achieve. Between perfect and no correspondence, there is a continuum without clear-cut "low", "medium", or "high" states. This wording demonstrates a relative difference or approximate range on the continuum. It is crucial to remember that fidelity is an objective concept that describes the degree of correspondence without judgment. Higher interaction fidelity is not necessarily better, more desirable, more effective, or more immersive but merely implies a closer match to the reference. On the other hand, aspects of fidelity often determine the success of a simulation. To ensure effective and economical planning, interaction designers and VR developers should reflect on what kind of fidelity is essential for the use case.

All aspects of fidelity can be assessed objectively and subjectively depending on the point of view. When applying standardized metrics for reproducible and indisputable measures to describe fidelity, we can objectively determine and verify the exactness of the interaction's match with the reference interaction. It might make sense to assess some aspects subjectively, such as perception or experience. Currently, we do not have the means to determine every aspect objectively. We may achieve this level of technical feasibility in the future, potentially even for experiential fidelity with sufficiently sophisticated brain-computer interfaces.

The model comprises the user and the VR system, which includes input devices, the computer as the processing unit with data and models, and output devices. Figure 2.1 illustrates the connections between these components with arrows, indicating the translation from software to hardware (such as the rendering from the simulation to the output devices) or, vice versa, from physical to intangible information (such as from the system output to the user's mind through perception). In an iterative process, we refined the labels and definitions of the components by engaging in discussions with research peers, in teaching practice, and at conferences. Table 2.1 lists the proposed terms and definitions.

We can consider the model vertically and distinguish aspects of *input fidelity* (right side) and *output fidelity* (left side). We can also consider the model horizontally and distinguish aspects of fidelity that concern the *user* (upper part) and those concerning *system fidelity* (lower part). As a whole, all aspects of the model define overall interaction fidelity. The single components can be further broken down as needed (e.g., interaction fidelity → simulation fidelity → presentational fidelity → 3D model → skin texture → height map → resolution). In the scope of **P1** (IntFi Model), we detail conceivable subcomponents through examples, not comprehensively. Only *simulation fidelity* is further divided into four subcomponents due to its structural complexity.

We conducted semi-structured, interactive expert interviews with 14 VR researchers and practitioners via Zoom or in person to improve and validate the model. After introducing the model, we asked the experts to apply it to an example project of theirs with a broad thematic diversity. The conversations lasted 63 minutes on average. We performed a thematic analysis of the 14.75 hours of material to process and structure the findings. The sample included three members of the IEEE VGTC VR Academy and an ACM Distinguished Scientist. The ten researchers in our sample had an average citation count of 7,043 and an average h-index of 35 as of February 14, 2024. The h-index ranged from 11 to 69 and reflects the spectrum of our selection that includes young scientists publishing with a focus close to the IntFi Model and experts with decades of experience in VR research.

Table 2.1: The definitions of the fidelity components and the categorizing sets of components.

Aspect	is defined as the degree of exactness with which...	Depends on
Action Fidelity	user actions resemble those of the reference interaction.	User
Detection Fidelity	input devices detect the user's actions.	System
Transfer Fidelity	virtual actions, derived from the input measurements, resemble the user's actions of the reference interaction.	System
Simulation Fidelity	a virtual environment resembles the characteristics of the reference interaction's world and adequately reacts to the user's actions.	System
Rendering Fidelity	the output content generated by the computer resembles what would be presented to the user in the reference interaction.	System
Display Fidelity	the output devices reproduce the physical stimuli presented to the user in the reference interaction.	System
Perceptual Fidelity	the user's perception of the physical stimuli created by the system resembles how the user would perceive the reference interaction.	User
Experiential Fidelity	the user's experience of the simulated interaction resembles how the user would experience the reference interaction.	User

Set of aspects	is defined as the degree of exactness with which...	Includes
Input Fidelity	the virtual actions generated from the user's input resemble the user actions of the reference interaction.	Action, Detection, Transfer
Output Fidelity	the system output is generated and perceived as the user would perceive it in the reference interaction.	Rendering, Display, Perceptual
System Fidelity	the system reproduces the world of the reference interaction reacting to the user.	Detection, Transfer, Simulation, Rendering, Display
Interaction Fidelity	**reference interactions are reproduced.**	**All aspects**

Overall, the interviewed experts described the model as “useful,” “comprehensible,” “sound,” and “a meaningful contribution.” Each of our interview partners appreciated the IntFi Model as interesting and helpful. However, there was also reasoned criticism, opposing views, and a need for clarifying discussions. We report on the rich qualitative data in the paper along the identified themes of *concept criticism*, *terminology criticism*, *applicability criticism*, *application strategies*, *application traps*, *patterns*, and *contributions*.

VR developers, designers, researchers, and managers can use the proposed meta-model to understand, compare, hypothesize, and teach the factors that define fidelity in VR. We demonstrate in **P1** (IntFi Model) how the model can be applied as a tool using three detailed examples. Additionally, the model is applied by analyzing the upcoming studies included in this thesis. With such analyses, the IntFi Model can help identify relationships and dependencies of the single components. If we identify such patterns by connecting the dots of different studies, we can gain a deeper understanding and derive strategies for effective and predictable VR experiences. To set the stage, we introduce five conceivable patterns of typical relationships in Section 7.1 of **P1** (IntFi Model), such as the Bottleneck Pattern and the Irrevocable-Loss Pattern. I will identify some of these patterns in the upcoming analyses of the other publications.

Furthermore, the paper provides substantially more details and considerations beyond this summary. Regarding the single components, we delineate the aspects, detail their characteristics, specify their requirements for maximum fidelity, illustrate how different levels of fidelity could be designed, and reference specialized frameworks or similar definitions. We present guidelines for applying the IntFi Model purposefully and effectively, share application strategies for different use cases, and describe application traps to be cautious of. We provide in-depth examples of analyses using the model and

Figure 2.3: An overview of the central contributions of the IntFi Model. *Icons from flaticon.com (modified).*

illustrative materials for teaching and using the model in teams. Additionally, we discuss the conceptual framework in a broader context and derive foundational research opportunities. Figure 2.3 provides an overview of the work's contributions. I include similar figures at the end of each paper summary to visualize the contributions and findings at a glance.

With its eight components, the IntFi Model illustrates what design decisions influence the fidelity and characteristics of VR systems, enabling immersive and realistic experiences. Zoomed out, the model provides a holistic framework integrating all elements of the interaction. When zooming in, further specialized models are needed to investigate the underlying complexity of a component. Breaking down this complexity would exceed the scope of this meta-model and requires the combined efforts of the VR research community. **P1** (IntFi Model) instead serves as a signpost to specialized models. One of them is the Haptic Fidelity Framework. It zooms in and further untangles one of the output modalities, as haptics affect rendering, display, and perceptual fidelity.

2.2 | The Haptic Fidelity Framework

This section is based on publication **P2 (Haptic Fidelity)**:

> Thomas Muender, Michael Bonfert, Anke Reinschluessel, Rainer Malaka, and Tanja Döring. 2022. **Haptic Fidelity Framework: Defining the Factors of Realistic Haptic Feedback for Virtual Reality.** In Proceedings of the 2022 Conference on Human Factors in Computing Systems (CHI '22), Association for Computing Machinery.

My contribution to this work: I contributed to the research idea, the theoretical considerations in terms of structure and definitions, the illustrative examples, and the conception of the scoring approach (*conceptualization*). I further contributed to the literature review, the validation method and process (*data curation*), and the writing and editing of the manuscript (*writing – draft, review & editing*).

We created the Haptic Fidelity Framework to provide a comprehensive and technology-agnostic framework for haptics. It can be used to assess any haptic feedback system in detail. As haptic experiences are highly subjective, it can be challenging to understand the underlying processes consistently. Therefore, we established objective criteria that rely on perceptual psychology and technological properties.

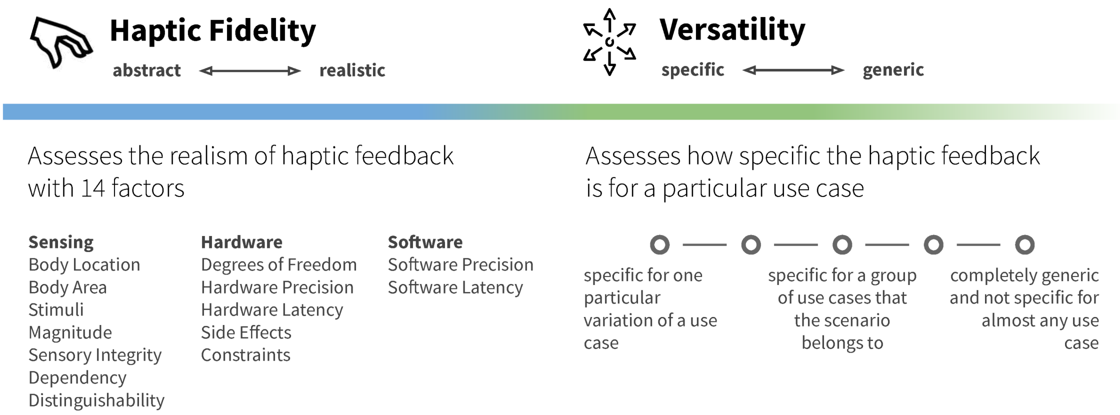


Figure 2.4: Overview of the framework's two dimensions, Haptic Fidelity and Versatility, with their respective factors and categories.

The characteristics of haptic feedback can be determined along two dimensions of the framework. First, *Haptic Fidelity* describes the degree of exactness with which the haptic stimuli and perceptions correspond to the reference interaction. It is determined by 14 factors that can be grouped into *Sensing*, *Hardware*, and *Software*, as shown in Figure 2.4. These groups approximately align with the IntFi Model's components of output fidelity: rendering fidelity (Software), display fidelity (Hardware), and perceptual fidelity (Sensing). Second, the dimension of *Versatility* describes how specific the haptic feedback is for a particular reference interaction. It has five ordinal categories, from being specific to one particular variation of a use case to being completely generic.

Haptic feedback contributes substantially to making a virtual world feel tangible and more realistic. Various devices have been devised to achieve this, including standard VR controllers with vibration feedback, steering wheels, and custom-built controllers that simulate shape, weight, and forces, as well as electric muscle stimulation (EMS), actuated robot arms, drones, and passive haptic props. These systems vary significantly regarding their technologies, goals, and approaches to providing feedback. This diversity makes it challenging to compare systems, obtain comprehensive evaluation results, and inform the design of new systems. Current frameworks either focus on the user experience (Kim and Schneider, 2020), target specific types of feedback, such as vibration feedback (Obrist et al., 2013), or propose physical measures that only apply to certain types of feedback systems (Samur, 2012). The factors defined in our framework can be adapted to any concept of haptic display. For example, the factor *Magnitude* can describe how adequate the intensity is of a force-feedback glove's exerted forces, of a controller's vibration strength, of an electrical muscle stimulation pulse, or of a proxy's weight.

In contrast to the wording of the IntFi Model, the framework originally refers to an "intended use case" instead of a reference interaction. Furthermore, the framework employs a spectrum ranging from abstract to realistic instead of low to maximum fidelity, because reality was the only reference frame considered. In this thesis, I stick to the IntFi Model's terminology for consistency.

The Haptic Fidelity factors can be assessed quantitatively to improve system comparability, resulting in an overall score between 0 and 4. We provide an expert tool[2] to rate the factors and automatically calculate the score. Three example evaluations of implementations for the reference interaction of grasping a mug are located within the two dimensions in Figure 2.5. We validated our framework and its scoring by applying it to 154 haptic feedback conditions from 38 research papers with user studies on haptics. We then compared the Haptic Fidelity score to the users' perceived realism from self-reports. The analysis showed a strong correlation, which confirms that systems with a high fidelity score in our framework are perceived as more realistic by users.

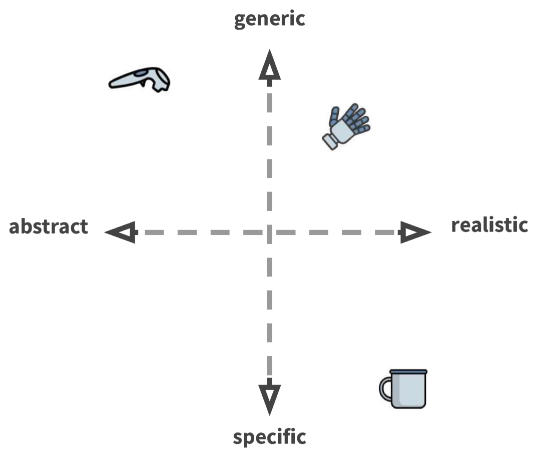


Figure 2.5: Classification of three examples in the framework's dimensions for providing haptic feedback for a virtual mug: a standard VR controller with vibration, a force-feedback glove, and a real mug as a passive haptic prop.

Furthermore, the analysis revealed a strong negative correlation between haptic fidelity and versatility, demonstrating a trade-off between the two dimensions in current haptic systems. The more realistic a device can provide haptic feedback, the more restricted in applicable use cases it is. Consequently, there is a lack of universal haptic interfaces that can generate feedback with high realism. This finding was only possible through systematically assessing many studies using the same theoretically sound framework in a meta-analysis. The included studies covered a wide variety of feedback modalities and device types, ensuring high external validity and generalizability. The complete analysis and materials are available in the paper's supplemental materials.

The framework enables researchers and designers of haptic feedback systems to anticipate the fidelity and versatility of a system's haptic feedback, hypothesize evaluation

[2] http://haptic-fidelity-framework.com

results, compare different systems or their variations, decide which solution fits a use case best, and understand the influencing factors of haptic feedback. We provide detailed examples of how to apply the framework in analyzing haptic systems. Figure 2.6 provides an overview of the work's contributions.

Similar to this framework's differentiated assessment of haptic fidelity, other components of the IntFi Model with their various modalities can be decomposed in the same granular way. Besides the Haptic Fidelity Framework, several other models and taxonomies assist in particularized assessments of other subcomponents, as indicated in Table 2 of **P1** (IntFi Model). In the following chapters, I will evaluate the presented studies in terms of fidelity based on the IntFi Model and the Haptic Fidelity Framework. After summarizing the publications, I will identify the respective reference interactions and analyze crucial aspects of interaction fidelity to interpret and connect the empirical findings.

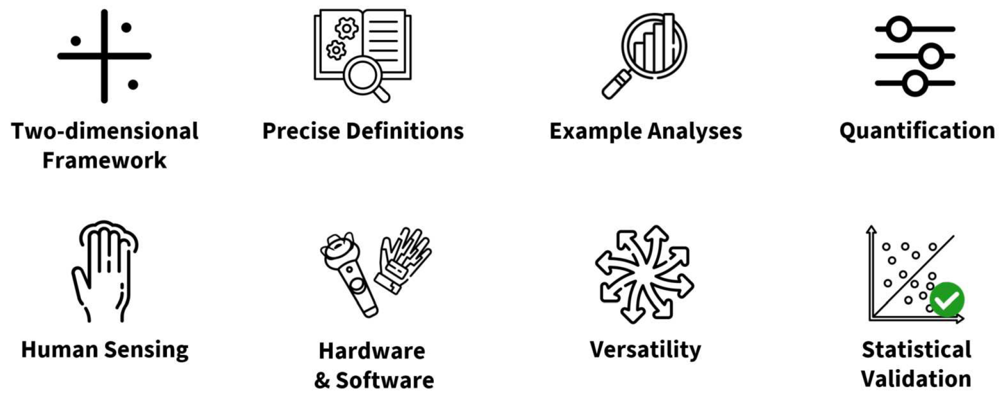


Figure 2.6: An overview of the central contributions of the Haptic Fidelity Framework. *Icons from flaticon.com (modified).*

3

Questionnaires in VR Research

As VR technology evolved over the last decades, established research methods have been adapted to investigate interactions with these systems. Typically, subjective mid- and post-experience measures are collected via self-reports in questionnaires (Lazar et al., 2017). Conventionally, study participants fill out paper or computer-based forms. However, when evaluating VR experiences, they first must remove the headset and change the domain from virtual to physical reality to access the questionnaire. This often leads to temporal disorientation and a loss of sense of control (Knibbe et al., 2018). Additionally, the participants are torn out of their environment shortly before they are supposed to assess it. Accordingly, questionnaire results are likely biased to a degree that is difficult to quantify and varies from case to case. Such undetermined bias is highly problematic for many types of research and evaluations.

VR allows integrating questionnaires directly into the environment, which we refer to as inVRQs in contrast to outVRQs. This offers the opportunity to stay closer to the ongoing simulation than out-of-VR research setups. Since no context change occurs, a break in presence (BIP) can be reduced (Jerald, 2016; Putze et al., 2020). This is especially important for time-sensitive measurements that must be realized as soon as possible after the exposure, such as simulator sickness (Keshavarz and Hecht, 2011) or presence (Skarbez et al., 2017).

At the same time, it is crucial to consider and report how questionnaires are embedded in VR to foster adequate interpretation of study outcomes and replicability. Schwind et al. (2019) observed higher consistency of self-reported presence using inVRQs, with no significant difference of means. The authors highlight that the effects of using questionnaires in VR are unclear, pointing out that the commonly applied measures were not validated for VR studies.

In this chapter, I address the second research question about adapting questionnaires to be used inside VR as investigated in publication **P3** (in-VR Questionnaires).

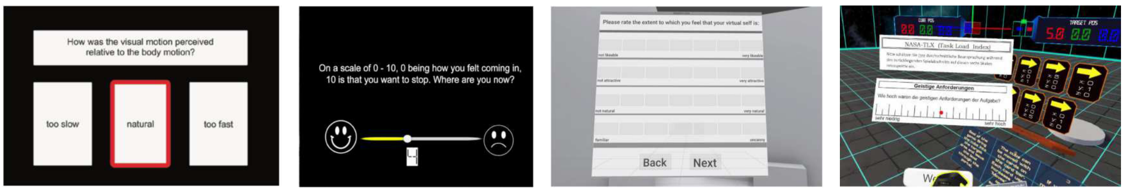
3.1 | Self-Reporting in VR

This section is based on publication **P3 (in-VR Questionnaires)**:

> Dmitry Alexandrovsky, Susanne Putze, Michael Bonfert, Sebastian Höffner, Pitt Michelmann, Dirk Wenig, Rainer Malaka, and Jan David Smeddinck. 2020. **Examining Design Choices of Questionnaires in VR User Studies.** In Proceedings of the 2020 Conference on Human Factors in Computing Systems (CHI '20), Association for Computing Machinery.

My contribution to this work: I contributed to the systematic literature review (*investigation*), to the research idea and prototype design (*conceptualization*) and study design (*methodology*) of the main user study, as well as to the writing and revision of the manuscript (*writing – draft, review & editing*).

To gain a better understanding of current practices, considerations, and recommendable design principles, we present four contributions in this research: (1) a literature review of 123 publications on VR user studies, (2) a survey with 74 VR experts, (3) a preliminary design study to compare different presentation and interaction methods, and (4) a user study (n=38) evaluating our refined inVRQ tool, comparing it to a laptop-based solution. This investigation focuses on quantitative self-reports, while qualitative responses and objective measurements are out of scope.



(a) Kang et al. (2018) **(b)** Cao et al. (2018) **(c)** Schwind et al. (2018) **(d)** Oberdörfer et al. (2019)

(e) Fernandes and Feiner (2016) **(f)** Schwind et al. (2019) **(g)** Wienrich et al. (2018)

Figure 3.1: Examples of in-VR questionnaire implementations: (a) and (b) present a head-anchored questionnaire, (c)–(f) use a world-anchored questionnaire, and (g) presents the questionnaire anchored to the body.

As a first step, a literature review provided an overview of questionnaire practices in current HCI research in the VR domain. We evaluated 123 publications from ACM CHI, ACM CHI Play, ACM VRST, IEEE VR, and IEEE 3DUI from the years 2016 to 2019, hence, between the release of the first HTC Vive and the submission of this paper. The findings show that most papers did not report how the questionnaires in their user studies were administered. From our selection, 15 papers describe using inVRQs in combination with outVRQs (12) or solely (3). In screenshots and descriptions, we identified different implementations in terms of input modality (pointing, trackpad, oral), devices (VR controller, gamepad, hand tracking), and canvas anchoring (world, body, head). Some examples are shown in Figure 3.1. Our analysis shows how uncommon the reporting and usage of inVRQs are and that the custom-built solutions differ significantly.

Secondly, we conducted an online survey with VR researchers on their research practices as well as experiences of and attitudes towards inVRQs. We received 67 valid responses. In line with our literature review, the survey found a lack of common questionnaire standards for VR user studies. Most experts had a positive attitude toward integrating them into the VE. However, some experts demonstrated a strong opinion against inVRQs. They explained it with technical challenges, implementation efforts, fear of response bias, and participant overload.

Lastly, we conducted two user studies investigating the design and usability of inVRQs from the users' perspectives. In a preliminary design study (n=10), we tested different interaction modalities (controller pointing vs. trackpad) and canvas anchoring

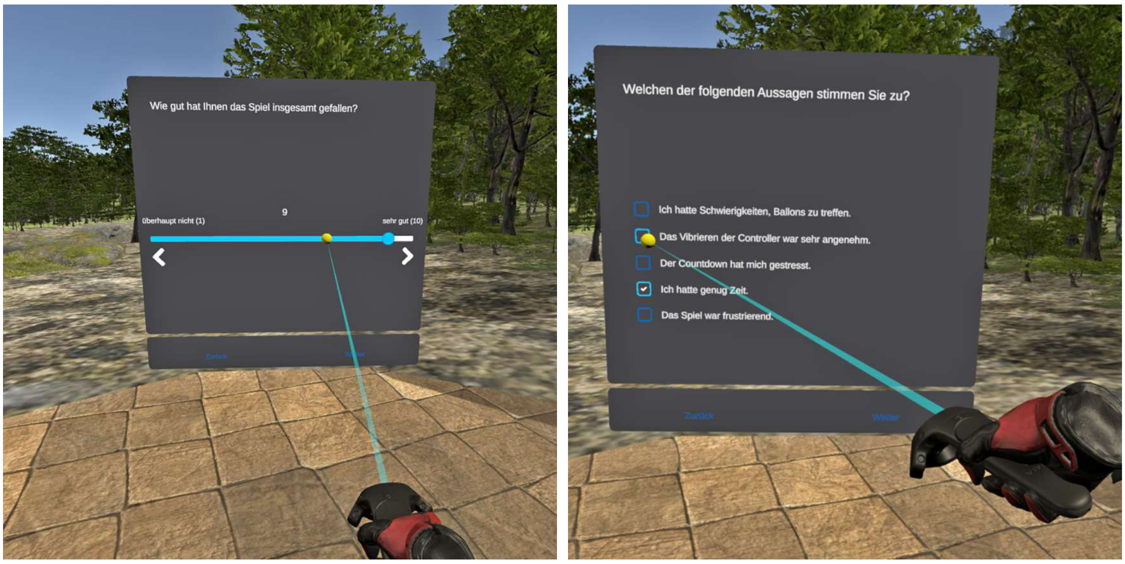


Figure 3.2: Screenshots of the inVRQ prototype evaluated in the usability study. *Images generated with Unity Editor®.*

(world-anchored vs. hand-anchored) with participants experienced in game design and VR. We measured completion times, System Usability Scale ratings (Brooke, 1996), and user preferences. We interviewed the participants at the end of the experiment. The results show clear benefits of world anchoring and pointing with a ray cast.

Consequently, we evaluated a refined implementation with world anchoring and pointing selection, as shown in Figure 3.2, in a within-subject usability study (n=38), comparing it to an outVRQ presented on a laptop with an external mouse. Most users (82%) preferred filling out the questionnaires in VR. Due to the unreliable responsiveness of the inVRQ tool, its usability was rated significantly lower than the outVRQ but with overall high UMUX usability scores (Finstad, 2010). There was no difference in the reported task load according to the NASA TLX (Hart, 1986) but higher physical demand and effort, which can be attributed to the standing position and mid-air controller pointing in VR. In contrast, participants sat and rested their hands on the table for the outVRQs. The data show no differences in presence ratings, contradicting the suspected response bias. Furthermore, the tools took a comparable amount of time. Based on our findings, we advocate for applying inVRQs in VR user studies. We demonstrated that world-anchoring with controller pointing provides a usable and efficient interaction design that participants prefer over conventional outVRQ solutions. Figure 3.3 provides an overview of the work's contributions and findings.

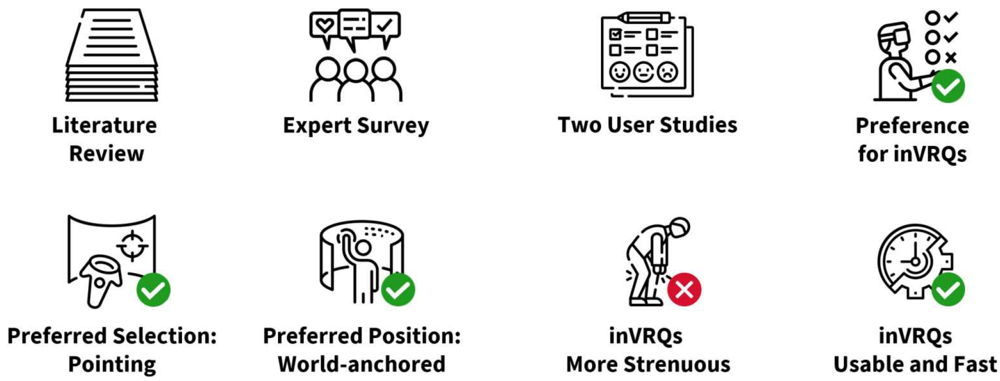


Figure 3.3: An overview of the central contributions and findings of the study. *Icons from flaticon.com (modified).*

3.2 | Fidelity of Questionnaires in VR Research

In response to the second research question, I now discuss **P3** (in-VR Questionnaires) in terms of fidelity.

RQ2: How can conventional user study questionnaires be transferred to VR with high fidelity?

To evaluate VR experiences, this established method of self-reporting on a laptop or paper was typically adopted—but not adapted. After the VR experience, participants were usually asked to leave the VE and answer a questionnaire back in reality. However, by integrating the questionnaires into VR, participants can stay in the same environment as the experience. This can reduce the break in presence that occurs when exiting VR (Jerald, 2016; Putze et al., 2020) and ensure a smoother experience for participants by eliminating the need to take off and put on equipment repeatedly. But if transferring questionnaires to VR was trivial, it would probably have been done for decades in VR research. Still, only 12% of the studies used inVRQs when we conducted this research, according to our literature analysis.

To replicate something virtually, we emphasize in the IntFi Model how important it is to identify the appropriate reference interaction. Virtually simulating filling out questionnaires seems to have a straightforward reference interaction. But imagine the difficulties of a VR user holding a virtual pencil and ticking boxes of a questionnaire on virtual paper. For example, Oberdörfer et al. (2019) represented the NASA TLX in VR as if on paper, as shown in Figure 3.1 (d). While this realization resembles the conventional real-life activity more closely, there is no benefit to the outcome of self-reporting. Therefore, some aspects of the IntFi Model, such as action or output fidelity, are negligible in this use case. On the other hand, striving for high fidelity to paper- or web-based interactions might result in significant usability issues for the participants. The fear of low usability and user frustration were among the top reasons against using inVRQs mentioned in our expert survey.

To adapt the methodological standard in VR studies, we should be careful not to copy the tools mindlessly and instead examine how VR questionnaires can be designed with high usability. Therefore, the more suitable reference interaction is merely selecting answers to visually presented questions. To accomplish this, we need to focus on transfer fidelity. Based on our literature review and expert survey, we evaluated different interaction techniques for answering inVRQs in two user studies. We decided to use handheld controllers for user input because they are commonly used in many current VR systems and provide high detection fidelity.

Our findings demonstrate how administering questionnaires in VR can be realized efficiently and with high usability. Participants prefer this solution over conventional outVRQs. Our proposed interaction design features world-anchoring of the presented questionnaire and pointing with a controller to select. This is one possible implementation to obtain participant responses but certainly not the only viable approach with high user satisfaction. In **P1** (IntFi Model), we highlight that high fidelity does not necessarily result in high usability, presence, or other benefits. Sometimes, high fidelity can be detrimental. It is crucial to reflect on which elements of the reference interaction must closely resemble for a successful simulation. In the examples of inVRQs, usability should be prioritized over interaction fidelity.

The studies in this paper focused on controller interactions and primarily investigated questionnaire anchoring and selection modality. To facilitate answering questionnaires without controllers, such as with hand tracking, force-feedback gloves, voice input, or brain interfaces, there is a need for examining further inVRQ interaction designs. External validity in VR research would benefit from better methodological standardization and easy-to-deploy toolkits, maximizing the comparability of study results. As our tested tool supports only sliders, radio lists, radio grids, and checklists, the research should be extended to further self-reporting elements, such as open-ended questions or dropdowns.

Our main user study found a high consistency in user responses collected within and outside VR. The means and variances of the presence measures were similar in both conditions. This is a reassuring indicator that the proposed inVRQs may attain comparable data to previously used procedures outside VR but should be confirmed in further studies with substantially larger statistical power. If the assessments differ, it is difficult to decide which outcome can be considered "valid." Eventually, we may need to rely on new standardized questionnaires designed for and validated in VR as the new gold standard in VR user research.

4

Object Manipulation in VR

In daily life, we frequently perform intricate tasks with our hands. Consider holding a pen in your hand, gently pressing its tip onto the paper, and shaping letters with fine movements orchestrated by all fingers. It requires a great deal of skill to make this work. This level of control over a pen depends on the dexterous abilities of our hand movements, the reliable response of the tool, the detailed haptic feedback we receive while performing the task, and our prior experiences.

When people first experience VR, they often check if objects behave like those in the real world. However, this is typically not the case in current consumer applications. It is challenging to virtually replicate the described complexity of object manipulation with high realism. Picking up virtual objects can be easily achieved by approximating a possible hand pose around it and attaching it to the virtual hand. Rudimentary haptic feedback is provided by the controller's contact forces or vibrations. While this implementation is technically simple, it is often not realistic.

However, certain use cases require true-to-life interactions for handling virtual objects, such as in medical training, manufacturing, telesurgery, digital marketing, or practicing motor skills. But how can we simulate authentic agency over and behavior of digital entities without physical manifestation? Due to our hands' dexterity and sensory richness, it is currently unfeasible to attain one universal interface that enables every type of object interaction. To approach a system that perfectly resembles manipulating and perceiving real objects, we can break down the complexity of the challenge and explore solutions to individual aspects of the problem. These solutions may address an object's shape, texture, temperature, adhesiveness, and the possibility of squeezing, throwing, stacking, and swinging it realistically.

This chapter presents two approaches for enhancing the realism of object handling in VR interfaces: varying the applied grip strength and perceiving virtual weight. With the publications **P4** (Variable Grip) and **P5** (Triggermuscle), I address the third research question concerning fidelity and object manipulation.

4.1 | Using Grip as Input for Rotational Control

This section is based on publication **P4 (Variable Grip)**:

Michael Bonfert, Maiko Hübinger, and Rainer Malaka. 2023. **Challenges of Controlling the Rotation of Virtual Objects with Variable Grip Using Force-Feedback Gloves.** Frontiers in Virtual Reality 4.

My contribution to this work: I contributed the research idea (*conceptualization*) and wrote the first draft of the manuscript (*writing – draft*). I contributed equally to the design of the interaction technique (*conceptualization*), study design (*methodology*), statistical analysis (*formal analysis*), the revision of the manuscript (*writing – review & editing*), and the figures (*visualization*) as MH. I made minor contributions to the prototype development (*software*). I supervised the project (*supervision*).

When we hold a virtual object and move or turn our hand, current VR systems often transfer the movement and rotation identically to the object. This limits the possible orientations of the object to how far we can revolve or contort our hand. Beyond or instead of this extrinsic movement of turning the entire hand, we would control a real object with intrinsic movements, i.e., use our fingers to adjust how we hold it within our hand (Elliott and Connolly, 1984). We can impact a held object through the pressure we apply to it. Our grip determines how the skin and the object's surface interact. For example, when holding an object with gentle pressure, we can let it slip through our fingers without dropping it.

In this work, we expand on an interaction technique proposed by Bonfert et al. (2019), which considers the grip applied to a held object using controller-based systems.

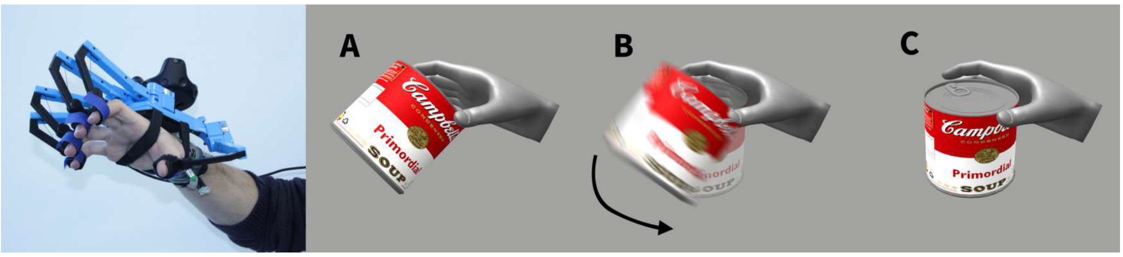


Figure 4.1: A user changes the orientation of a virtual can. **Left**: The SenseGlove DK1 is attached to the user's fingers, providing force feedback along the can's shape. **Right: A)** The user applies a firm grip. Thus, the rotation of the can is fixed. **B)** The can swings downwards due to gravity when the user loosens the grip by reducing the finger pressure. **C)** The can stays in level.

By varying the grip strength, the object can be held loosely, allowing it to rotate freely between the fingers, or it can be grasped firmly, which transfers the hand's rotation directly to the object, as illustrated in Figure 4.1. This grip variability affords control over more rotational degrees of freedom and can help avoid clutching (Zhai, 1996). The affordances of grip variability influence our everyday tasks in various situations—often subconsciously—for example, when moving a full glass, using a screwdriver, screwing in a light bulb, or fidgeting with a pen. We rely on the object's inertia or the pull of gravity to change its rotation. Virtually reproducing these natural grip-based manipulations could increase interaction fidelity. Beyond more realistic interactions, a variable grip can compensate for inadequate object orientations due to the initial grasp, such as for virtual tools.

For controller-based systems, the interaction technique showed advantages in terms of user satisfaction, intuitiveness, and realism with a slight decrease in the task load. The abstract button actions must be mapped to the intended virtual actions of controlling an object, which was found to increase mental demand (Bonfert et al., 2019). The idea was implemented for the Valve Index controllers in a follow-up study by Pedersen et al. (2023). Pressure applied to the controller's handle is interpreted as direct input for the grip strength applied to the object. This removed the need for mapping, which increases input fidelity. Again, an evaluation showed that the interactions with variable grip were considered more realistic, slightly slower, and with no difference in the NASA TLX (Hart, 1986) ratings. Only higher usability could not be confirmed.

As a follow-up, we realized the interaction technique with the SenseGlove DK1 VR gloves providing force feedback to the fingertips for increased action fidelity. With soft finger pressure, the object can rotate with three degrees of freedom around an anchor point in the middle between the fingertips of the thumb and index finger. More than 80% pressure locked the rotation to the hand's movement, visualized with a colored bar above the hand avatar. In a within-subject user study, 21 participants did pick-and-place tasks inspired by the study design by Bonfert et al. (2019). We compared performance, presence, task load, and agency between *fixed* and *variable grip* with 756 virtual cans, books, and milk cartons, which are shown in Figure 4.2.

The study results showed a poorer performance with the grip variability. The object placement was slightly less accurate, the study participants needed 40% longer, and they needed considerably more grasping attempts to move an object. This aligns with previous research on controller-based systems (Bonfert et al., 2019; Pedersen et al., 2023). In contrast to the outcome of this experiment, grip variability with controllers was reported to be more intuitive, easier to control, and more satisfactory (Bonfert et al., 2019), as well as more realistic (Bonfert et al., 2019; Pedersen et al., 2023), compared to the

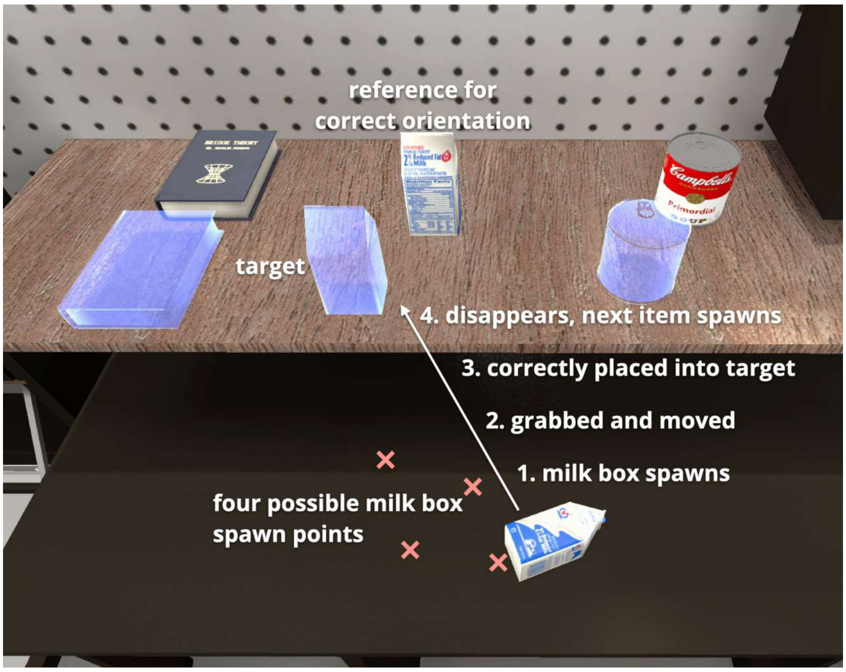


Figure 4.2: One of the two pick-and-place tasks in the experiment. Participants moved objects of different types from a spawning point to a target position in a given orientation. *Image generated with Unity Editor®*.

condition with a fixed grip. However, when using a force-feedback glove with a variable grip, participants reported higher mental load and frustration, inferior perceived performance, and less agency over the object behavior.

As possible explanations for these unexpected results, we elaborate on the general difficulties of handling objects with the SenseGlove DK1 in publication **P4** (Variable Grip). While unexpected and inadequate force feedback is already challenging with the fixed-grip baseline, a variable grip is even more confusing. Another reason might be the novice sample: 18 out of 21 participants have never used finger tracking for object manipulation. Due to the steep learning curve, participants assumed they could have performed better with considerably more practice and might then find it helpful. Moreover, the unsteady grasp of the variable grip mode caused more items to be accidentally dropped by the participants, which naturally resulted in longer handling times. Figure 4.3 provides an overview of the work's contributions and findings.

Overall, the presented study has not shown that mapping the finger pressure with force-feedback gloves to the grip strength of handling virtual objects would, in principle, be an undesirable solution. We have not yet demonstrated the potential we anticipate when using hardware with more accurate actuation, sophisticated force vector estimation of individual fingers, and additional cutaneous feedback.

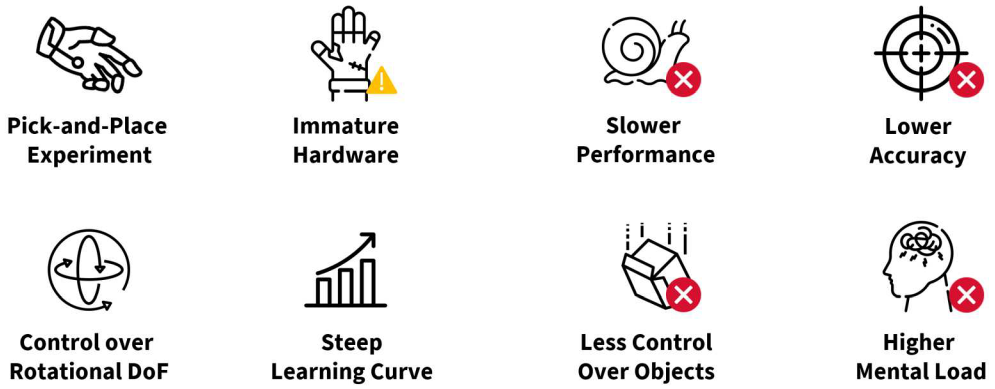


Figure 4.3: An overview of the central contributions and findings of the Variable Grip study. *Icons from flaticon.com (modified).*

4.2 | Creating a Sense of Weight with *Triggermuscle*

This section is based on publication **P5 (Triggermuscle)**:

Carolin Stellmacher, Michael Bonfert, Ernst Kruijff, and Johannes Schöning. 2022. **Triggermuscle: Exploring Weight Perception for Virtual Reality Through Adaptive Trigger Resistance in a Haptic VR Controller.** Frontiers in Virtual Reality 2.

My contribution to this work: I contributed to the discussion of the research idea (*conceptualization*), the study design (*methodology*), the literature review, and to writing and editing the manuscript (*writing – draft, review, & editing*). I supervised the project with JS (*supervision*).

When holding real objects, our muscles, tendons, and skin receptors sense the gravitational pull, informing us about their weight. In contrast, picking up virtual objects in VR with conventional VR controllers or hand tracking cannot give users any weight cues. Unless the system features a world-anchored haptic interface (e.g., a robot arm) or repurposes physical matter (e.g., props or liquids), no physical forces that resemble weight are rendered. In many cases, only visual cues are ambiguous and insufficient in conveying weight information. Conventional VR controllers are lightweight, quick to set up, and offer various input possibilities but can only create vibration feedback.

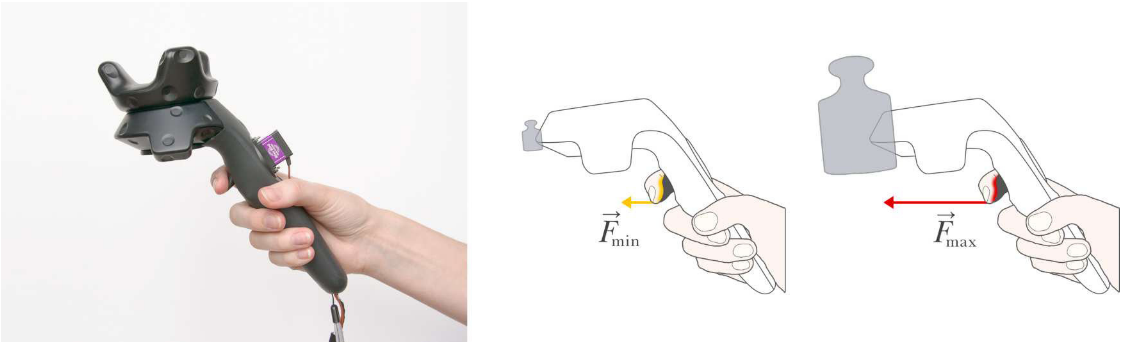


Figure 4.4: Our haptic controller *Triggermuscle* simulates the weight of virtual objects in VR by changing the trigger resistance. **(left)** The spring mechanism for a dynamic adjustment is built into the casing of an HTC Vive controller. An HTC Vive tracker is mounted to the top to enable spatial tracking. **(right)** Triggermuscle allows continuous regulation of the trigger resistance according to the weight of the grabbed virtual object.

But how can we provide haptic stimuli for an object's weight while maintaining the benefits of controllers? The human perception of weight cues informs our proposed approach. To grasp an object, the human brain incorporates haptic and visual cues (Gordon et al., 1991; Loomis and Lederman, 1986) and previous lifting experiences (Van Polanen and Davare, 2015) to predict an object's weight and scale finger forces accordingly. When lifting the object, we adjust the grip force according to proprioceptive and cutaneous stimuli (Brodie and Ross, 1984; McCloskey, 1974), resulting in a direct relationship between physical weight and applied finger forces. Previous research found that pulling the trigger to varying degrees can be interpreted by users as exerting variable grip forces on virtual objects (Bonfert et al., 2019).

Therefore, we suggest utilizing the trigger for weight cues as a standard button of any consumer VR controller. We can provide dynamic haptic feedback through existing hardware components by varying the trigger's resistance, usually constant by default. As a result, users need to scale their index finger force according to the virtual weight of a held object, as indicated in Figure 4.4. The heavier it is, the more finger force must be exerted onto the trigger. While it requires the pressure of more than one finger to grasp real objects, we map this exertion to the index finger, commonly used for grasping with VR controllers. To demonstrate and evaluate this novel principle, we devised two haptic controllers with different spring mechanisms built into the casing of an HTC Vive controller: a preliminary prototype and an improved controller called *Triggermuscle*. We evaluated both devices in user studies to investigate users' ability to discriminate different levels of resistance and the effect on weight perception in VR.

To test our initial prototype with adaptive trigger resistance, we performed a psychophysical experiment (n=9) following the method of constant stimuli with a two-

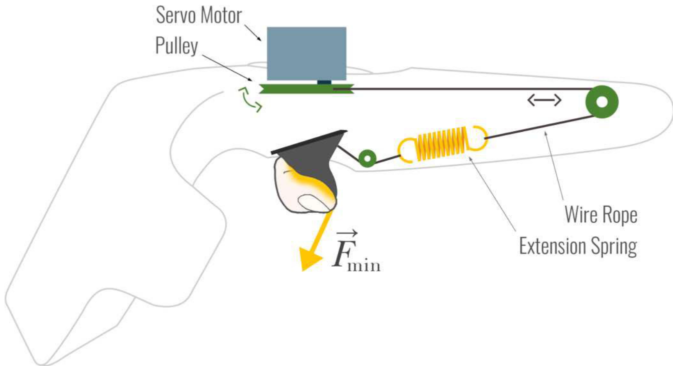


Figure 4.5: Schematic illustration of Triggermuscle's spring mechanism utilizing an extension spring and a servo motor.

alternative forced choice (2AFC) paradigm (Jones and Tan, 2013). Participants were asked to indicate which of the two presented stimuli felt heavier to them. By comparing combinations of different trigger resistance levels, we calculated the just-noticeable difference (JND), the point of subjective equality (PSE), and the Weber Fraction (WF), which is the ratio of the JND to the initial spring tension. We observed surprisingly mixed results. Some participants identified almost all stimulus differences correctly—even subtle changes—while others sensed no change and performed around the guess rate. Only data from three participants qualified for JND computation.

We refined the hardware, software, and experiment design according to our preliminary findings. The improved device, Triggermuscle, features the spring mechanism illustrated in Figure 4.5 with a 134% larger force range (4.29 to 16.36 N) and a versatile form factor suitable for various controller shapes. We tested Triggermuscle in a 2AFC experiment (n=21) in which users lifted virtual boxes and compared the perceived weight rendered with different trigger resistances. Again, the results demonstrate a highly diverse perception of trigger resistance, similar to the outcome of the first study. We found close-to-perfect stimuli recognition, close-to-guessing behavior, and only four data sets that exhibit the expected decrease in discrimination ability with decreasing resistance intensity. Overall, three out of five people successfully differentiated Triggermuscle's haptic stimuli. In interviews at the end of the experiment, only one-third of participants mentioned the adaptive trigger resistance as the cue for identifying the heavier box. Others reported interpreting vibrations from the servo motor's adjustment, visual cues, or sound as the basis for their weight interpretation.

Considering the successful stimulus discrimination and the interpretation of the varying trigger resistance as a weight cue, our studies confirm the validity of our approach of sensory substitution with haptic stimuli at the index finger. However, given that some participants could not perceive the change in trigger resistance or did not associate it with weight, there seem to be significant individual differences in perception and interpretation, questioning the mechanism's general applicability in the proposed implementation. Six Triggermuscle users seem to have subconsciously interpreted the adaptive trigger resistance correctly, although, in the interview, they could not determine why. One explanation could be a visual dominance (Hecht and Reiner, 2009; Posner et al., 1976) despite identical-looking boxes, as participants' visual attention might have distracted from the haptic experience. This effect has been previously reported in the context of pseudo-haptics, haptic retargeting, and rendering shape or texture in VR. Furthermore, the side effects of servo vibrations might have been a haptic confounding factor. With a more sophisticated technical implementation, our proposed approach for weight simulation in VR could be a simple and effective upgrade of controllers with various form factors. Our paper discusses limitations in hardware design and promising research opportunities for weight perception through hand-held VR controllers. Figure 4.6 provides an overview of the work's contributions and findings.

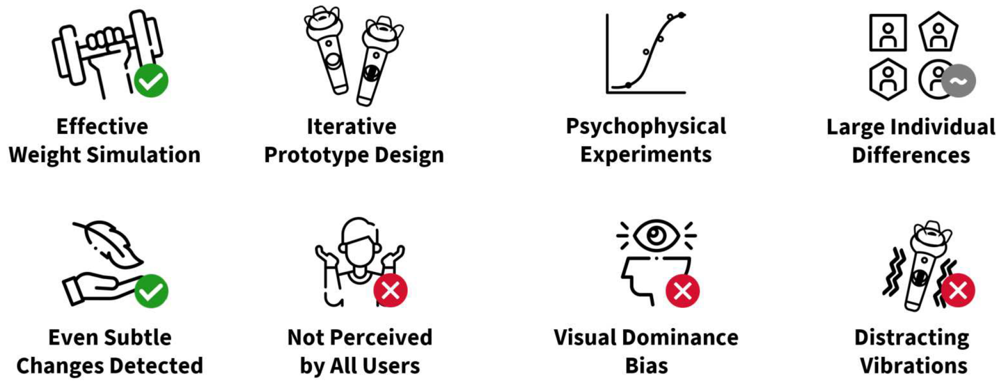


Figure 4.6: An overview of the central contributions and findings of the two studies. *Icons from flaticon.com (modified).*

4.3 | Fidelity of Object Manipulation in VR

Regarding the third research question, I next analyze the fidelity of the interactions from **P4** (Variable Grip) and **P5** (Triggermuscle) based on the theory from Chapter 2.

RQ3: What are the effects of varying interaction fidelity for manipulating virtual objects on perception and performance?

Realism of Object Handling with Variable Grip

The study from **P4** (Variable Grip) builds on an interaction technique first implemented and evaluated for a controller-based system (Bonfert et al., 2019). As the basis for further analysis, we first need to inspect the original approach from a fidelity perspective. We can identify a clear reference interaction: When handling real objects with our hands, we control the grip strength applied to the object's surface. The friction ensures a safe grasp without crushing the object. Softening the grip allows a controlled rotation of the object within the hand—a way to manipulate objects that has been impossible in VR before. The tested interaction technique replicates this manual ability by mapping the input of controller buttons to the rotational freedom of the object.

Applying the IntFi Model, how can we describe aspects of interaction fidelity for this approach? In the real world, we adjust our finger forces pressing against the object to vary the grip and, thus, how freely it can move. In contrast, in the original controller implementation, the user presses a button to adjust the grip strength. Such an abstract means of input has low action fidelity. Similarly, output fidelity is low regarding haptics since the haptic cues from the object's surface are missing. The interaction technique only works because transfer fidelity is high due to the close correspondence of the virtual grip actions to the real-life equivalent. As a result, the user study found high usability, intuitiveness, controllability, comfort, and perceived realism, with only slightly reduced performance compared to always-firm grip. Users took more time handling objects with a variable grip and reported a higher mental load as the mapping requires cognitive effort.

A follow-up study by Pedersen et al. (2023) evaluated a system that requires no button mapping. As I proposed in the original publication (Bonfert et al., 2019), they used the Index controllers, which detect the pressure of the user's hand squeezing the handle. With this, the grip forces of all fingers are used as input. This can be interpreted as more direct control over a held object, leading to higher action fidelity. However, their study showed ambiguous results with potentially higher completion times and an unclear effect on task load. Users rated the perceived realism using variable grip as higher.

In the next step, we avoided abstract mapping and aimed to increase the technique's action fidelity further while raising output fidelity. If the control over the object and its perception are closer to reality, users need to think less about the interaction and can handle objects more intuitively. To achieve this, we implemented a system with force-feedback gloves. We can assess high action fidelity, as the hand encloses the object and exerts forces onto its surface, interpreted as grip strength, and increased output fidelity, as the glove provides dynamic kinesthetic forces simulating the resistance of the object.

But instead of the anticipated usability boost, the outcome was worse than in the original study, with low usability and inferior performance. How can we explain the drop in user satisfaction despite closer-to-reality interaction? We stepped into the uncanny valley of VR interactions, as first coined by McMahan et al. (2016). This phenomenon is included as one of the fidelity patterns in the IntFi Model. It describes a non-linear correlation of interface quality and interaction fidelity that predicts good performance with low-fidelity systems and even better performance with high-fidelity systems but a drop in performance for medium fidelity. We elaborate on this pattern in Section 7.1. of **P1** (IntFi Model). In the following, I discuss several relevant issues of the tested exoskeleton system that might explain the outcome.

One of the major issues was the insufficient haptic feedback. Due to high input fidelity and the output fidelity assumed to be high because of the kinesthetic feedback, users might expect maximum output fidelity, although this is not the case. The Haptic Fidelity Framework helps identify the issue. The foundational factor *Stimuli* describes "the degree to which the same haptic receptors of the user are involved" as in the reference interaction. Table 2 of **P2** (Haptic Fidelity) lists the haptic stimuli associated with sensing various physical properties—in this case, for perceiving contact forces and friction. Although the gloves render constant kinesthetic forces, which are required for perceiving contact forces, they render no tangential forces resembling skin stretches, which are always involved in the perception of friction (Augurelle et al., 2003; Cadoret and Smith, 1996). If these cutaneous cues are missing, as studies have shown using local anesthesia, people drop real objects or use an overly powerful grip because their mental model of the held object's physical properties is insufficiently informed (Augurelle et al., 2003; Westling and Johansson, 1984).

Furthermore, other possibly involved stimuli, such as pressure changes or vibrations, are not rendered. Therefore, haptic fidelity is lower than users might initially assume, with a Haptic Fidelity score of 1.90 (from 0 to 4).[1] The system lacks the necessary cutaneous cues for intuitively controlling how an object slides between the fingers.

[1]Use Factor Code pDZRApEw on haptic-fidelity-framework.com to check all factor ratings.

Further possible reasons for the findings include the general difficulties in handling objects with the SenseGlove DK1 as an early development prototype. Considering the simple task, we measured relatively high NASA TLX scores and slow placement even in the baseline condition with a simple pick-and-place task. According to interview feedback, many users did not feel confident using the glove in any condition. This could be due to severe limitations of the force feedback mechanism as it depends on the angle at which the fingers press against the surface. The more oblique a finger touches the surface, the higher the chance of no or unexpected feedback. While this is already confusing with the fixed-grip baseline, it is even more confusing with a variable grip. It makes it hard to control the applied grip as it depends on the unreliable resistance.

The next step in follow-up studies could be to increase output fidelity further. For variable grip strength, supporting tactile cues beyond kinesthetic forces must be considered in multi-finger object manipulation. Shear forces, friction, slip, or contact forces have been shown to improve reproducing real-world haptic experiences (Girard et al., 2016; Kim et al., 2022; Salazar et al., 2020; Whitmire et al., 2018). Also, the impact of their combination with kinesthetic feedback from gloves would be insightful. The ideal prototype for exploring the benefits of natural grip-based object manipulation would render detailed tactile cues, rely on complex physics calculations comprising the individual pressure of all fingers, afford dynamic and continuous grip, and avoid interpenetration of the object and the hand. Besides imitating grip strength, future research could investigate other physical factors in object manipulation with force-feedback gloves. For example, in relation to **P5** (Triggermuscle), the principle of variable trigger resistance correlating with the object's virtual weight could be applied to the finger resistance of a force-feedback glove. The heavier the object, the more force must be exerted by the user to lift it with the glove.

Realism of Weight Perception with Adaptive Trigger Resistance

The reference interaction that the interface with adjustable trigger resistance resembles is picking up and holding objects of different weights. The standard example used in **P1** (IntFi Model) is also used to demonstrate Triggermuscle: grasping an apple. There, the weight sensation of holding an apple is compared to the perceived weight of holding a strawberry and a pineapple. Similar to the variable grip, the mechanism behind Triggermuscle aims to convey a haptic experience from the virtual object, although it has no mass and thus experiences no gravitational pull. While the variable grip can give control over the object's rotation utilizing its virtual weight, Triggermuscle is designed

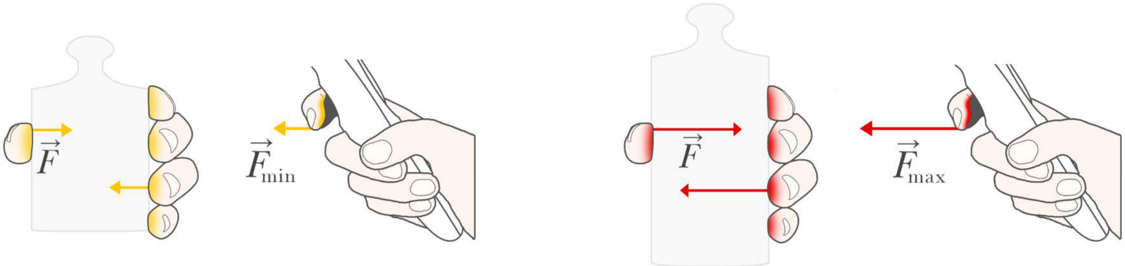


Figure 4.7: The mapping of grip forces when holding an object to the force of the trigger resistance against the index finger. *Figure from* ***P5*** *(Triggermuscle).*

to convey the sensation of weight itself. I analyze this replication using the IntFi Model and the Haptic Fidelity Framework.

When we hold an object, the grip forces of pressing the fingers against its surface must increase with its weight for a safe grasp. In Triggermuscle, this is mapped to the force that the index finger must exert on the trigger, as illustrated in Figure 4.7. With this transfer function, transfer fidelity is high, as the virtual actions closely correspond to the real actions. Action fidelity is relatively low because the grasping action is reduced to just one finger with a small movement and restricted by the hand-held controller. Still, it is higher than for conventional grasping with controllers, as the required grasping forces must scale with the object's mass. Similarly, output fidelity is higher due to the adaptive kinesthetic feedback for heavier objects.

For most users, the increase in action, transfer, and output fidelity successfully induced a sense of weight, thus increasing experiential fidelity. This is remarkable, as the sense of weight occurs despite low perceptual fidelity due to the lack of actual gravitational forces.

Nevertheless, some users have not linked the haptic cues to the object's weight or have not even noticed the changing trigger resistance. To understand this discrepancy better, we again consult **P2** (Haptic Fidelity). According to Table 2, perceiving weight always requires sustained pressure and possibly pressure changes, constant forces, and force changes, all of which are displayed to the index finger by Triggermuscle. However, skin stretches are also possibly involved in weight perception but are not rendered by Triggermuscle. Moreover, vibrations are never involved in sensing weight, yet Triggermuscle's servo motor generates noticeable vibrations as a side effect. Overall, we can assess the system's Haptic Fidelity score with only 0.42 (from 0 to 4).[2] Among other factors discussed in **P5** (Triggermuscle), these missing and misleading haptic stimuli might have confused some users or distracted them from the intended stimuli. According-

[2]Use Factor Code `0qaJA5og` on http://haptic-fidelity-framework.com to check all factor ratings.

ing to the Loss-Propagation Pattern introduced in Section 7.1 of **P1** (IntFi Model), the low display fidelity inherits these limitations to the subsequent fidelity components in the loop. Nevertheless, interaction fidelity was high enough to successfully convey a sense of weight to most users despite shortcomings in haptic fidelity.

As the next steps in enhancing adaptive trigger resistance, the system design could be refined to reduce confounding and add missing haptic cues. Testing the system outside VR with blindfolded users could help identify the influence of visual dominance. Follow-up studies could also explore explanations for the observed individual differences in stimulus perception, such as finger sensitivity, fine motor skills, gaming experience, frequent practice of delicate hand tasks such as playing an instrument, or regular mindfulness practice.

Furthermore, combining the kinesthetic feedback with other techniques of weight simulation might amplify the desired effect. For example, a follow-up study by Stellmacher et al. (2023) investigated the combination of Triggermuscle with pseudo haptics. The authors modified the control-display (C/D) ratio during lifting in addition to the adaptive trigger resistance. With the combined techniques, users were more sensitive and faster in identifying weight differences, which demonstrates how integrating different approaches can facilitate closer-to-reality object manipulation.

5

Embodiment in VR

We can experience and affect VEs in various ways with different modalities. In the previous chapter, we looked at how interaction techniques and haptic feedback can increase the fidelity of manipulating objects and how this can influence the user experience. However, the proposed solutions require specific devices that are often expensive and specific in their applications, thus being less common. Therefore, it is interesting to consider possibilities for facilitating high-fidelity interactions solely based on visual rendering, as most VR experiences rely on visuals. How can we effectively increase the realism of VR interactions by only adjusting what the users see? Which steps have an impact commensurate with the implementation effort?

In particular, this chapter focuses on the visual appearance of virtual humans, both of oneself and others. The perception of our body and appearance strongly impacts how present and capable we feel in a VE (Schultze, 2010). Also, the appearance of other virtual humans strongly affects how we perceive them and influences social interactions (Bailenson et al., 2005; Zibrek and McDonnell, 2019), be it avatars representing other people or system-controlled agents. The possibilities for visualizing virtual humans are broad. A virtual human can be represented by a simple shape indicating a position or anthropomorphic. It can look abstract, cartoony, or photorealistic. The entire body or only parts of the body can be visible. For a user's avatar, the whole body can be precisely tracked and displayed, or it can be only partially tracked, which requires extrapolating body movements, such as with inverse kinematics. It can also be invisible or anything between the mentioned characteristics. A recent large-scale study with 2,509 knowledge workers investigated attitudes towards differing realism of avatars in a work context (Phadnis et al., 2023). Participants preferred high realism for known and unknown colleagues but sometimes perceived the most realistic virtual humans as uncanny. The acceptability of avatars correlated with their fidelity. Non-realistic avatars were perceived as playful and less professional, which was experienced as unsuitable for work in some countries.

An avatar's appearance affects not only the people seeing it but also the person it embodies. Similar to being in a biological body, we can have the feeling of having and controlling a virtual body. This sense of embodiment has been described as a construct with three components: the senses of self-location, agency, and body ownership (Kilteni et al., 2012). These components can be identified in the example of the rubber hand illusion experiment (Botvinick and Cohen, 1998). In this experiment, the participants can see a rubber hand right next to their own hand being touched. The real hand is covered and invisible. The participants experience the touch of the rubber hand as if it were their own hand and attribute it to their bodies—a sense of body ownership. They have the impression that their hand is in the location where they can see the rubber hand—a sense of self-location. When the experimenter applies physical violence to the rubber hand, the participants withdraw their hands in horror. As the rubber hand does not respond to their movement, their sense of agency over the rubber hand breaks. This experiment has been replicated in various variations (Riemer et al., 2019), also in VR with virtual hands (IJsselsteijn et al., 2006; Yuan and Steed, 2010) with similar outcomes. The visual appearance of the virtual hand affects how strong the embodiment is (Pyasik et al., 2020).

With this chapter, I address the fourth research question about fidelity and aspects of embodiment, as elaborated on in publications **P6** (Kicking), **P7** (VRisbee), **P8** (Agent Embodiment), and **P9** (VR Meetings).

5.1 | Foot Visibility in VR Sports

This section is based on publication **P6 (Kicking)**:

> Michael Bonfert, Stella Lemke, Robert Porzel, and Rainer Malaka. 2022. **Kicking in Virtual Reality: The Influence of Foot Visibility on the Shooting Experience and Accuracy.** In Proceedings of the 2022 Conference on Virtual Reality and 3D User Interfaces (IEEE VR '22), IEEE Computer Society.

My contribution to this work: I contributed the research question (*conceptualization*) and statistical analysis (*data curation* and *formal analysis*). I contributed the majority of the study design (*methodology*). I wrote and edited the majority of the manuscript (*writing – draft, review, & editing*) and created all the figures and presentation material (*visualization*). A made minor contributions to the prototype development (*software*). I supervised the project (*supervision*).

Figure 5.1: (left) A side view of the player's visualized foot before kicking the virtual ball. (right) The targets to shoot. The second window from the left was just hit and broke. The fourth is flashing as an indicator to be shot next. The player's perspective is further away and centered. *Images generated with Unity Editor®.*

Hand-based VR interactions gained more attention in research, industrial applications, and consumer systems than foot-based interactions. However, the feet play a vital and versatile role in our lives far beyond walking. Especially in some sports, they are essential. In soccer—or football in the UK—we use our feet to kick the ball, pass it to other players, and score goals. When shooting at or towards a target, professional soccer players look neither at the ball nor their feet. Instead, they lock their view onto the target when taking the shot to improve their aiming. In the real world, they can feel haptic feedback from the ball while dribbling and shooting. When simulating soccer games in VR, the impact of the virtual ball cannot be perceived on the player's foot unless some form of force feedback is rendered, such as with active actuators in the shoe (Rovers and Essen, 2006) or ball props (Bozgeyikli and Bozgeyikli, 2019).

In this work, we investigated precise target shooting by kicking a virtual soccer ball in VR using only a Vive Tracker attached to the foot as an input device and providing no haptic feedback. We present an empirical study that explores the effects of a visual representation of the player's virtual foot on the interaction with the ball, as corresponding visualizations of virtual hands have been shown to improve the feeling of presence and body ownership (Canales et al., 2019; Grubert et al., 2018; Ma and Hommel, 2015). The between-group experiment examines if the visibility of the player's virtual foot (H1) improves the performance in terms of accuracy and required time, (H2) reduces the task load, (H3) enhances the presence in the virtual environment, (H4) increases the subjective control over the ball, and (H5) enhances the perceived body ownership.

To evaluate the effects of visualization on foot interaction in VR, we conducted a user study (N=28) with a kicking task inspired by soccer penalty shootouts. The sample included inexperienced and advanced soccer players. The players were asked to kick a soccer ball from a stationary position at a 11.22 m distance and hit eight targets. In the between-groups experiment, we compared the shooting performance, task load, presence, ball control, and body ownership between two conditions: kicking a ball with a visible foot (+F) *versus* with an invisible foot (–F). Players with foot visualization could see the virtual soccer shoe and sock from Figure 5.1. A Vive Tracker was attached to the right foot with elastic bands. The left foot was neither tracked nor virtually displayed. After 30 practicing shots, the participants had to hit eight targets in a given order, with the ball respawning in the same location. Questionnaires and a short interview followed the task.

The data showed that the accuracy improved significantly by showing the foot. On average, players with invisible feet needed 58% more attempts to hit a target—up to 21 shots per target. The best-performing player needed less than two shots on average and could see their foot. With a visible foot, the players were over one minute faster; however, this difference is not statistically significant. On the other hand, the perceived body ownership was rated significantly better by players who could see the foot. Players without visualization were in considerable disagreement, as evidenced by a large variance. We did not find a difference between the means of the ratings on ball control. However, the foot visibility caused a large variance in perceived control over the ball. As an explanation, we assume the varied calibration accuracy among the sample. Potential discrepancies were not so apparent for players with invisible feet, leading to unanimous ratings. However, for players who could see their shooting foot, any discrepancy became obvious. This might have led to either less ball control or, if the calibration was good, an improved agency over the ball. We also compared the perceived task load and sense of presence between the conditions and found no differences. Furthermore,

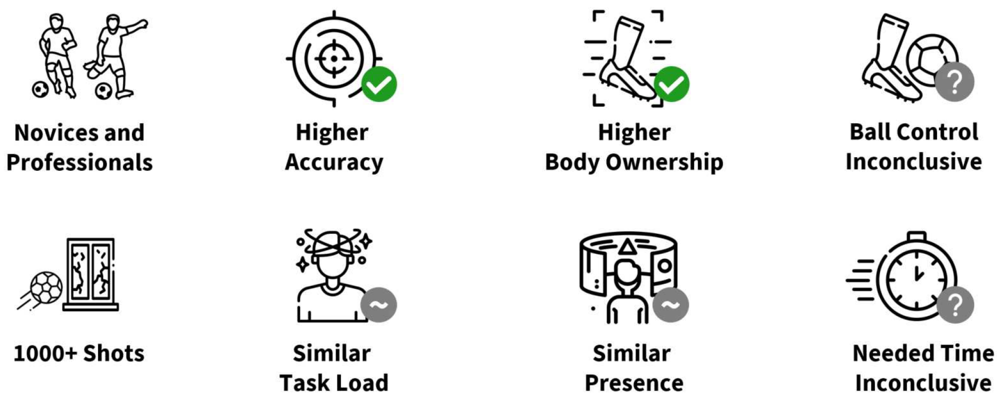


Figure 5.2: An overview of the study's contributions and findings indicating where we found significant differences, no differences, or inconclusive data. *Icons from flaticon.com (modified)*.

there was no measurable influence of the players' age, gender, or experience in soccer, VR, and gaming. Many of our participants shared how much fun they had during the experiment. They wanted to continue playing and practice their skills.

In conclusion, accuracy and body ownership were higher when showing a virtual foot, even independent of previous soccer experience. Although most players from the control condition said they did not miss seeing the foot, they needed many additional shots. One player only became aware of the foot visualization at the end of the session—still, it might have improved his aiming. From what we observed and learned in the interviews, we suspect effects on a subconscious level. The visual information allows continuous motor re-adjustments and a faster learning effect. These feedback loops seem especially plausible with the measured high body ownership. Figure 5.2 provides an overview of the work's contributions and findings.

5.2 | Hand Visibility in VR Sports

This section is based on publication **P7 (VRisbee)**:

Malte Borgwardt, Jonas Boueke, María Fernanda Sanabria, Michael Bonfert, and Robert Porzel. 2023. **VRisbee: How Hand Visibility Impacts Throwing Accuracy and Experience in Virtual Reality.** In Extended Abstracts of the 2023 Conference on Human Factors in Computing Systems (CHI EA '23), Association for Computing Machinery.

My contribution to this work: I contributed to the formulation of the research goals (*conceptualization*) and the study design (*methodology*). I made an equal contribution to the statistical analysis (*formal analysis*) as JB. I made a major contribution to writing and editing the manuscript (*writing – review & editing*). Together with RP, I supervised the project (*supervision*).

Building on the insights from the previous section that discussed the significance of visualizing feet for kicking in VR to enhance body ownership and accuracy in target shooting, we now turn to the role of hand visualization in sports. Previous research investigated skill acquisition and transfer of throwing in VR, such as with balls or darts, and found trade-offs in distance perception and aiming (Butkus and Ceponis, 2019; Mousavi et al., 2018; Tirp et al., 2015; Zindulka et al., 2020). Throwing a disk differs from other hand-based sports as it relies on different physical principles. In activities such as *Ulti-*

Figure 5.3: The prototype used for Frisbee throwing in the experiment. (left) A virtual hand holding a disk. (right) The view from the player's perspective on the bullseye target and the scoreboard next to it. *Images generated with Unity Editor®*.

mate Frisbee or *Disc Golf*, the players guide the disks with their hands on a flat trajectory stabilized by spin and lift forces that are highly dependent on the disk's orientation. In soccer, ball trajectory control is brief and limited to the kick's impact, whereas when throwing a Frisbee, prolonged disk interaction allows for more nuanced control over the disk's flight and involves richer tactile feedback while gripping and releasing the disk. Additionally, the player needs to consider the weight and shape of the disk as well as external factors such as wind through haptic stimuli. In current VR systems with controllers or hand tracking, these cues are unavailable to the player. The absence of tactile cues from touching the Frisbee, combined with its particular physical flight characteristics poses a unique challenge.

This raises the question of how the visibility of virtual hands affects disk throwing in VR. Studies found significant benefits of hand visibility for tasks like typing on a virtual keyboard (Grubert et al., 2018) or picking and placing objects (Argelaguet et al., 2016; Canales et al., 2019) as it has a positive influence on players' presence and body ownership in the virtual world. In this work, we investigated the impact of hand visibility on the performance and sense of embodiment when throwing a disk in VR along the following hypotheses: The visibility of the virtual hands (H1) improves the accuracy of throwing a virtual Frisbee, (H2) increases perceived presence, (H3) increases the sense of body ownership, and (H4) increases the subjective control over the virtual Frisbee.

In a user study, 29 participants were asked to pick up and throw a disk onto a round bullseye target 10 meters from the player with a diameter of 3 meters, as shown in Figure 5.3. The goal was to hit as close to the center as possible. In the between-group experiment, only half of the sample could see floating hands that were animated to a gripping pose when picking up a disk and opened up again when releasing it. The other group did not see a visual representation of their virtual hands. Five participants were left-handed. The simulation was implemented with Unity and the integrated XR Interaction Toolkit for the Meta Quest 1. The complex Frisbee flying physics is grounded in a numerical simulation that estimates the aerodynamic coefficients to match the trajectory of a real disk (Hubbard and Hummel, 8ftftft; Hummel, 8ftft2). We adjusted the virtual disk's translational velocity, angular velocity, and tilt parameters in Unity according to the estimated forces of the initial impulse, lift, drag, and torques. Every player had ten training shots and 2ftattempts to score as many points as possible in a standing position. After filling out the questionnaires, the session concluded with a short interview.

We observed enhancements in throwing accuracy, although with a small effect size only that may not be practically meaningful. The scatter pattern of disk impacts on the target is more dispersed on the horizontal axis. The analysis revealed distinct throwing patterns for left-handed and right-handed players. Right-handed throws demonstrated a rightward and upward skew. Given the movement of a natural backhand throw of a disk, this indicates that the release timing with a controller is difficult and that players tend to release the disk too late. There were no significant differences in the presence ratings except for the *realism* subscale, where a large effect size was observed. Differences in body ownership were not significant but borderline (p=.051). It is conceivable that with a larger sample, a possible effect might be measurable. With only a small effect on accuracy, this indicates that the visual focus while aiming lies mainly on the target and that the launch is controlled by proprioception or muscle memory rather than visual cues. Furthermore, the participants indicated a significantly improved subjective control over the disk with a large effect size. Moreover, the level of engagement was high, and numerous participants chose to continue playing after the experiment.

In conclusion, the subjective control over the disk and perceived realism were higher for throwing a Frisbee when seeing the virtual hands. The effects on accuracy and body

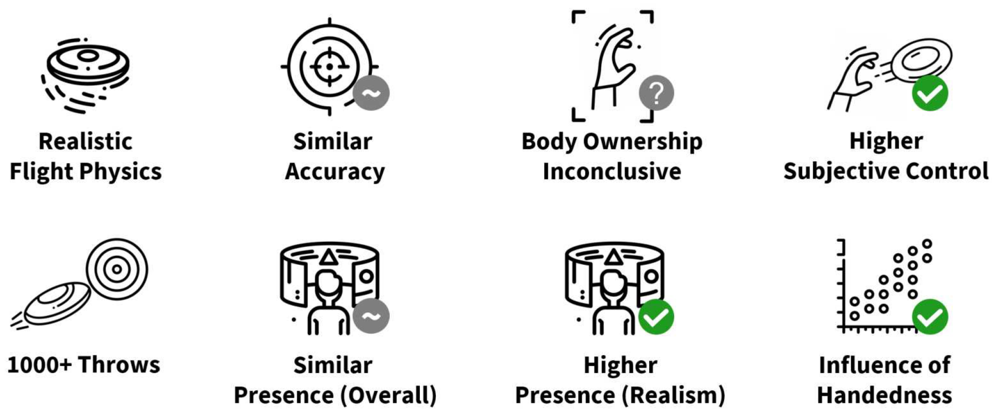


Figure 5.4: An overview of the study's contributions and findings, indicating where we found significant differences, no differences, or inconclusive data. *Icons from flaticon.com (modified).*

ownership are small and inconclusive. The study demonstrates that visible hands raise the feeling of agency over the virtual disk and support the user in secondary tasks, such as picking up the disk before throwing it. The results align with previous research on the challenging timing of releasing when throwing virtually (Zindulka et al., 2020). However, in contrast to the study in **P6** (Kicking) and a study on climbing in VR (Kosmalla et al., 2020), our findings showed that a visual representation of limbs does not considerably impact performance. Figure 5.4 provides an overview of the work's contributions and findings.

5.3 | Agent Visibility for Smart Displays

This section is based on publication **P8 (Agent Embodiment)**:

> Michael Bonfert, Nima Zargham, Florian Saade, Robert Porzel, and Rainer Malaka. 2021. **An Evaluation of Visual Embodiment for Voice Assistants on Smart Displays.** In Proceedings of the 2021 Conference on Conversational User Interfaces (CUI '21), Association for Computing Machinery.

My contribution to this work: I made equal contributions to the research idea (*conceptualization*) and the study design (*methodology*) as NZ and FS. I contributed the majority of the quantitative statistical analysis (*data curation* and *formal analysis*). I wrote and edited the majority of the manuscript (*writing – draft, review, & editing*) and created all figures and presentation material (*visualization*). I supervised the project together with NZ (*supervision*).

Smart displays are speakers with a touchscreen that provide users at home with a conversational and graphical user interface (UI). The user can talk to the integrated voice assistant and use the screen like a tablet. With the device's exterior, the system has a physical embodiment within the room, but the assisting agent commonly does not. The user can only hear but not see the agent despite the screen. In this work, we investigate the impact of embodying the agent on-screen to make it more present, signal availability, and visually convey human characteristics. Researchers have previously studied diverse types of embodiment for conversational agents (André, 2011; Isbister and Doyle, 2002; Lankes et al., 2007) as well as their visual attractiveness (Khan and De Angeli, 2009). Embodied virtual agents have become a natural extension of conversational interfaces by enriching the experience visually (Andrist et al., 2017; Cassell et al., 1999;

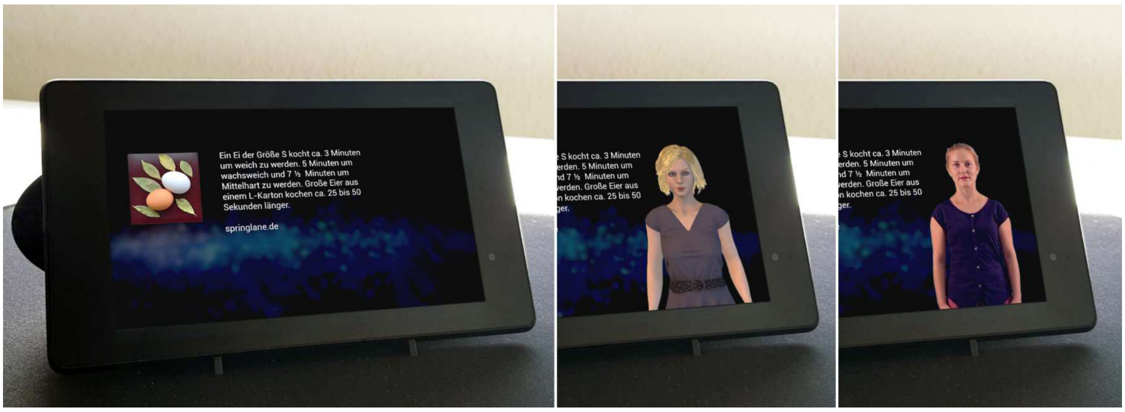


Figure 5.5: The three versions of the smart speaker prototype: (left) without agent visualization, (middle) with a digitally rendered agent, (right) with a photorealistic agent.

Nass et al., 1994; Wang et al., 2019). Today, smart displays present new opportunities for embodied conversational agents. Therefore, we present a user study investigating the impact of the agent's visual embodiment and the degree of visual realism on the user experience.

We conducted a Wizard-of-Oz experiment with a between-groups design in which the participants (N=60) interacted with one of the three prototype versions to complete a specified set of tasks. The activities represent a morning scenario and include a broad range of everyday commands (Kinsella and Mutchler, 2019). We designed three versions of a smart display: one with a disembodied agent (DEA), one with a digitally rendered, artificial embodied agent (AEA), and one with a prerecorded, photorealistic embodied agent (PEA) performed by a human actress. All versions had the same functionality and only differed in appearance, as shown in Figure 5.5. The agent was called "Joy" and spoke the local official language, German.

The quantitative results from the questionnaire ratings show that all conditions can result in comparably good user experiences. There were no significant differences between the conditions on any of the subscales. This indicates that agent embodiment does not impede the system's typical usage. The systems received medium to high ratings for their pragmatic qualities and attractiveness, and medium ratings for their hedonic qualities. After the experiment, we demonstrated the other two system versions to users. More than half of the 56 participants who created a ranking favored the photorealistic agent (51.8%). Only one-eighth of users would select the artificial agent (12.5%), and every third person preferred the version with a disembodied agent (35.7%). Only a few people expressed thankfulness toward the photorealistic agent, significantly less than in the other conditions.

The qualitative results show diverse opinions on the visual appearance of the agent. A pragmatic group of users preferred no embodiment as it would take up space, or a non-humanoid, abstract visualization to emphasize its artificiality. Others found a human-like appearance more natural and trustworthy. Many participants specified their preferred hair color, age, gender, and other characteristics. Some suggest more playful appearances, such as animals, fictional characters, celebrities, the user's self-avatar, a team of different experts, or even Microsoft Office's assistant *Clippy*. Furthermore, while some users liked that the agent's continuous on-screen presence indicated steady availability, the majority expressed unease about feeling observed or starred at. They would prefer the agent to disappear in idle mode or look distracted while not speaking.

Our work identifies critical design considerations for embodying voice assistant agents on smart displays to achieve higher user satisfaction. While the quantitative methods showed no significant differences in user experience, the interview results provided interesting considerations of the agent's appearance, customizability, on-screen presence while idle, and applying social conventions in the interaction. A third of the participants preferred the status quo of a disembodied agent, while the majority would like to see an agent that is either photorealistically humanoid or obviously non-humanoid.

Due to the conflicting preferences and strong opinions, we recommend allowing users to control the optional embodiment of the agent and its appearance. Moreover, we recommend hiding the agent between tasks to avoid social awkwardness and domestic intrusion. The reappearance can indicate that the system recognized the wake word and is listening to commands. Figure 5.6 provides an overview of the work's contributions and findings.

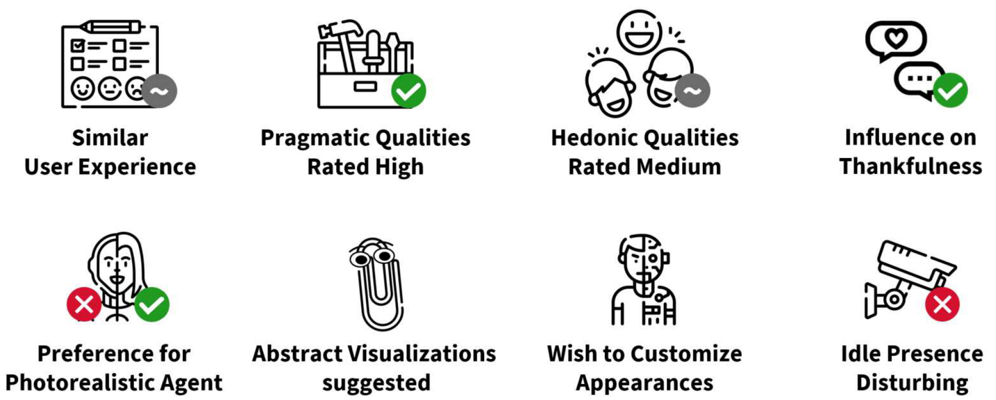


Figure 5.6: An overview of the study's contributions and results indicating quantitative findings and important design considerations. *Icons from flaticon.com (modified).*

5.4 | Peer Visibility for Meetings in VR

This section is based on publication **P9 (VR Meetings)**:

Michael Bonfert, Anke V. Reinschluessel, Susanne Putze, Yenchin Lai, Dmitry Alexandrovsky, Rainer Malaka, and Tanja Döring. 2023. **"Seeing the faces is so important"—Experiences from online team meetings on commercial virtual reality platforms.** Frontiers in Virtual Reality 3.

My contribution to this work: I had the leading role in contributing to the research idea (*conceptualization*) and study design (*methodology*). I made major contributions to the setup of the test environments (*software*), data acquisition (*data curation*), and qualitative data analysis (*formal analysis*). I created the majority of the figures (except the plots) and presentation material (*visualization*). I made a major contribution to the writing of the manuscript (*writing – draft, review, & editing*). I primarily coordinated the project (*project administration*).

After the previous publication examined how computer-controlled agents are perceived, we now turn to digital representations of real persons in remote collaboration. The Covid-19 pandemic required many people to work from the home office. Online meetings with video conferencing software suddenly became omnipresent. Although video calls provide many advantages, such as seeing meeting participants or allowing for screen sharing, they still yield limitations, such as restricted social interaction between participants. Social VR platforms might provide beneficial alternatives for online team meetings by gathering everyone in one virtual room. Previous studies showed positive psychological effects of social VR platforms (Barreda-Ángeles and Hartmann, 2021) and that group behaviors and emotional responses to it are largely similar to face-to-face encounters (Moustafa and Steed, 2018). Beyond the faithful reproduction of in-person meetings, research further explores the possibilities of VR for enhancing social encounters and collaboration outside the restrictions of reality (McVeigh-Schultz and Isbister, 2021; Slater and Sanchez-Vives, 2016).

To explore the potential of current off-the-shelf VR meeting software and gain first-hand insights into the advantages and drawbacks of authentic meetings in VR, we conducted the Digital Media Lab's regular team meetings in VR to compare the attendees' experiences with meetings on video conferencing platforms. Before the pandemic, the lab meetings were held in person. Between March 2020 and the beginning of this case

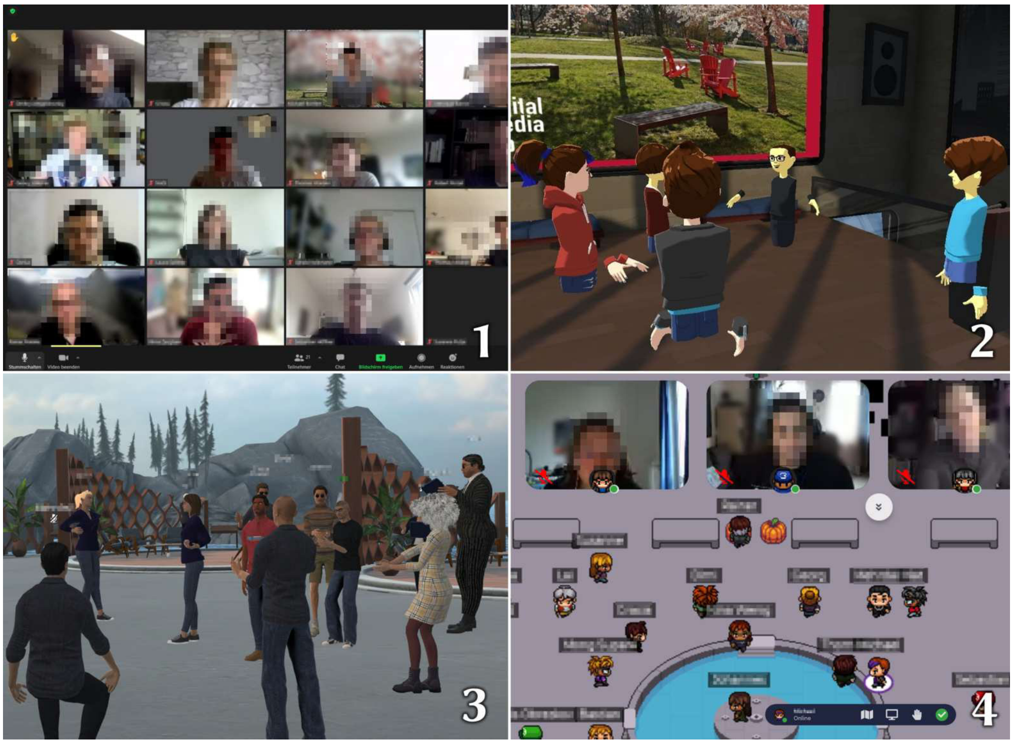


Figure 5.7: The four commercial platforms compared in our weekly team meetings: (1) Our status quo, Zoom, compared to the two virtual reality platforms (2) AltspaceVR and (3) Engage, as well as the hybrid (4) Gather Town combining video feeds with a 2D spatial environment. StarLeaf was also used but is not shown here as its interface is similar to Zoom.

study, they were on Zoom or StarLeaf. The shift to VR was intrinsically motivated, not by conducting this study. Therefore, the authors had the opportunity to evaluate the experiment independently with mixed methods, resulting in genuine insights from experiences in the wild. Previous studies explored attendees' experiences at professional social events in VR, e.g., at academic conferences and workshops (Erickson et al., 2011; Kirchner and Nordin Forsberg, 2021; Lahlou et al., 2021; Williamson et al., 2021), and group dynamics in social VR outside professional context (Moustafa and Steed, 2018; Scavarelli et al., 2021). However, in contrast to previous literature, this case study focuses on the participants' personal experiences with a heterogeneous sample during regular online team meetings in different mediums over an extended period with high external validity.

Over four months in 2020, we conducted twelve meetings on five platforms: seven meetings in VR and five on video conferencing or hybrid platforms for comparison.[1]

[1]Zoom: https://zoom.us; StarLeaf shut down in October 2022: https://en.wikipedia.org/wiki/StarLeaf; AltspaceVR shut down in March 2023: https://en.wikipedia.org/wiki/AltspaceVR; Engage: https://engagevr.io; Gather Town: https://www.gather.town

The tested platforms are shown in Figure 5.7. After each session, the participants were invited to complete a questionnaire, resulting in 8ftft responses with quantitative and qualitative data. As part of the lab team, the authors also attended the meetings, which enabled an additional autoethnographic perspective. Between 15 and 84 people attended the meetings. Twelve participants responded after at least ten meetings.

The quantitative data revealed a significantly better overall meeting experience on the video conferencing platforms than in VR. There was no difference in the ratings of the two tested VR platforms, AltspaceVR and Engage. The VR meeting experience was better for people with more previous VR experience. An exploratory factor analysis identified the three factors F1 *Involvement*, F8 *Co-Presence*, and F2 *Privacy* in our data. Users felt more involved and had higher co-presence with their colleagues in video calls than in VR meetings. There was no difference in the answers to the Privacy questions.

The rich qualitative results provide detailed explanations for the participants' preference for videoconferences along five themes. The users shared their impressions and opinions on spatial aspects (movement, proxemics, group dynamics, and spatial audio), the meeting atmosphere (co-presence, interpersonal interactions, professionalism, playfulness, and social norms), the expression of emotions (facial expressions, body language, and emojis), the meeting productivity (attention, distractions, presentations, secondary tasks, and tools), and their user needs (technical literacy, hardware discomfort, and privacy).

While users appreciated the free arrangement in the virtual worlds with benefits for natural turn-taking in group conversations and complex group dynamics, the VR meetings also caused more technical problems, induced cybersickness or discomfort for some attendees, and restricted secondary tasks. In contrast, videoconferences al-

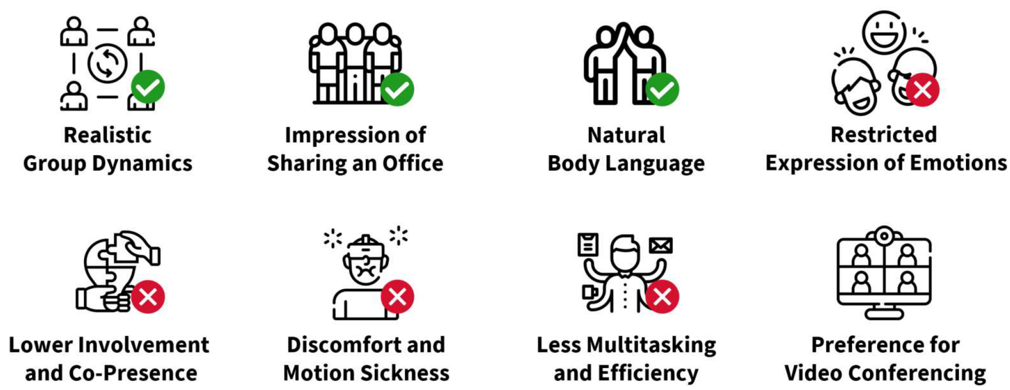


Figure 5.8: An overview of the study's contributions and findings indicating some of the benefits and shortcomings of the VR Lab meetings. *Icons from flaticon.com (modified).*

low seeing the faces and emotions of colleagues without technical preparation or nausea. We suspect that the strongest contributor to the higher co-presence ratings of video calls was the possibility of seeing each others' faces. Consequently, at the end of the experiment, most of the group argued for returning to desktop-based video conferencing.

Some benefits and challenges found in this study have been discussed in related literature. Still, the authentic in-the-wild setting of the intrinsically motivated exploration of how suitable current commercial platforms are under natural working conditions enabled us to bring practical issues into context and highlight critical research gaps. The paper outlines lessons learned with recommendations and links to prior research for each theme. Moreover, we analyze which challenges might be overcome soon with technical advancements and which require careful consideration of meeting format and technical aspects. Figure 5.8 provides an overview of the work's contributions and findings.

5.5 | Fidelity of Embodiment in VR

We now turn to the fourth research question. In this section, I analyze the interactions described in **P6** (Kicking), **P7** (VRisbee), **P8** (Agent Embodiment), and **P9** (VR Meetings) regarding their fidelity, on the foundation of the theory from Chapter 2.

RQ4: What are the effects of varying interaction fidelity on how the embodiment of others and oneself is perceived?

Realism of Foot and Hand Visibility for VR Sports

In the studies of **P6** (Kicking) and **P7** (VRisbee), the same experimental variable is manipulated: showing or hiding limbs of the user's self-avatar. For target shooting with a ball, the foot is visible or not, and for throwing a disk, the hand is. The reference interactions for this simulation are the real-life activities of kicking a ball or throwing a Frisbee, in which we can see our limbs. Therefore, the manipulated component from the IntFi Model is rendering fidelity, as the user's virtual limb model affects the virtual environment but is not visibly represented.

Since the experiments are so similar, we could assume comparable outcomes. Indeed, the study results point in similar directions, yet the detailed findings differ. Foot visibility substantially improved the performance and sense of embodiment, while hand visibility only affected these parameters statistically negligibly. On the other hand, the perceived control over the Frisbee was higher with visible hands, while the collected data on ball control was inconclusive. The condition with hand visibility was rated as more realistic on a subscale of the Presence Questionnaire (Witmer and Singer, 1998), while there was no difference on any of the subscales for kicking.

There are several possible explanations for why the group differences in player behavior and assessments differed between the experiments. A self-evident difference is the affected body part. The visibility of hands versus feet could have different impacts on the performance and the perception of the users' self-avatar. Another difference is the extent of the player's contact with the ball or disk. When kicking, the contact is short with a sudden impact, while there is extended contact with a prolonged impulse transfer when throwing. This involves different transfer functions but with high transfer fidelity in both cases due to the respective reference interactions. This implies a third significant difference for the conditions with invisible limbs. When holding the disk before releasing it, its visibility indicates the hand position as an extension of the virtual body. In contrast, the ball only serves as a reference point at the moment of impact—too late to correct the shot. This puts soccer without visible foot at a disadvantage and might explain the substantial performance advantage in the condition with foot representation.

Moreover, we observed the Bottleneck Pattern proposed in **P1** (IntFi Model) in the findings of the **P6** (Kicking) study. The foot–ball collision calculations proved imprecise for some players, resulting in poorly controllable shots, probably due to insufficient tracker calibrations. The low detection fidelity limited the physical simulation fidelity, which drastically constrained the player's experiential fidelity. According to the Bottleneck pattern, "experiential fidelity cannot be higher than a limiting key component, even if other aspects have much higher fidelity." Accordingly, affected players reported unrealistic and unsatisfactory ball responses independent of the experimental condition. In contrast, the **P7** (VRisbee) study employed hand controllers requiring no calibration and relied on more sophisticated physics calculations. This prevented the Bottleneck Pattern from occurring.

We can conclude that modifying interaction fidelity for motor-skill activities in VR sports significantly affects objective performance and subjective experience. Increased rendering fidelity by visually representing the virtual limbs with a co-located 3D model of a foot or a hand improves various factors in both studies. Visual feedback of the player's body parts can improve targeting accuracy—thus increasing action fidelity—and enhance subjective impressions of agency, body ownership, and realism—thus increasing experiential fidelity.

To better understand the role of haptic feedback in VR sports, replicating the presented studies would be insightful when using hand tracking or a force-feedback glove instead of controllers for disk throwing or with an actuated shoe for kicking. Furthermore, a longitudinal study could inform about potential learning effects as a more reliable muscle memory could be established with additional practice. To learn more about the impact of foot visualization on the factors yielding an inconclusive study outcome, we suggest investigating activities that focus visual attention on the feet, such as balancing or juggling the ball with the feet, and more complex soccer elements, like dribbling or passing the ball in a multi-user scenario. Furthermore, as throwing a ball in reality is twice as accurate as in VR (Zindulka et al., 2020), we suggest a comparison of real and virtual kicking to understand the unique characteristics of foot-eye coordination in VR. Controlling a virtual ball with the foot is more immediate than with a hand-held controller because there is no delay in releasing a trigger button, which reduces the control in throwing with the hand. This might lead to different findings between hands and feet. To complement the research on throwing in VR, it would be interesting to investigate how virtual catching compares to real catching. The findings of **P7** (VRisbee) suggest that hand visibility could have an even more significant influence on the catching performance.

Realism of Conversational Agents

The research in **P8** (Agent Embodiment) explores how users feel about the appearance of virtual agents. In contrast to the other publications included in this dissertation, this study is outside the VR domain. Instead, the users interact with virtual humans on a 2D smart display. Still, the IntFi Model can be applied with some precaution. The reference interaction to be simulated is a face-to-face conversation with a personal assistant whom the user asks for help to answer questions and perform tasks. The experiment compares three versions of the simulation: (DEA) the status quo of a disembodied agent whose voice is all that can be heard, which has minimum visual rendering fidelity due to the absence of a visible representation; (AEA) an artificial embodied agent whose voice and appearance is digitally rendered, which has low rendering fidelity due to the cartoony style and low simulation fidelity due to the simple agent behavior; and (PEA) a photorealistic embodied agent performed by a human actress, which has maximum rendering fidelity and relatively high simulation fidelity.

The findings show a clear preference for the agent with high realism. Most participants felt most comfortable with the photorealistic agent and would choose it as their favorite. One-third of users prefer the disembodied version, emphasizing the pragmatic priorities since the functionality is identical but without a superfluous virtual human. Only every eighth user preferred the artificial agent. The qualitative data indicates a strong uncanny valley effect of eerie agents that resemble a person—but not closely enough (Diel et al., 2021; Mori, 1970). MacDorman et al. (2009) believe that a computer-generated face is not necessarily eeriest when it looks nearly human, and argue that even abstract faces can look uncanny. The users who prefer a less realistic agent suggested more discernably artificial and abstract visualizations, which could be an attempt to underline the agent's artificiality while deliberately avoiding its eeriness.

Remarkably, the data showed an unexpected trend of less politeness towards the high-fidelity agent. We expected more courtesy in interactions with a more human-like assistant due to previous work showing skeptical opinions on whether a voice assistant is entitled to politeness at all (Bonfert et al., 2018). One explanation could be that the reactions by the actress are obviously recorded before the interaction; hence, the agent cannot rejoice in the expressed thankfulness, resulting in lower experiential fidelity. In contrast, the artificial and disembodied agents were rather experienced as a "live" artificial intelligence, which can be affected by and appreciate the user's politeness during runtime. Of course, behavioral simulation fidelity was, in fact, identical across all conditions, as no agent had the ability to rejoice.

Many users perceived the agent's continued presence between prompts as disturbing. This comes as no surprise if we recall the reference interaction. When talking to human assistants, it would be disconcerting if they hung around and stared at you while waiting for the next inquiry. This limit of behavioral simulation fidelity irritated users and impaired experiential fidelity.

In the future, virtual home assistants could be more common through MR interfaces than on smart screens. A follow-up study could investigate different options of agent embodiment for assistants who share the user's 3D space, such as their living room. An abstract or anthropomorphic visualization could lead to a different user experience for projections into the room than on a flat, small screen. In future work, it could also be helpful to differentiate between aspects of realism, such as visual and auditory rendering fidelity, as well as behavioral simulation fidelity in terms of animations or linguistic features.

Fidelity of Meeting Peers in VR

Lastly, we look through the lens of the IntFi Model to assess the fidelity of virtually embodied peers interacting with each other at VR meetings. As the weekly lab meetings initially took place in person in a large meeting room before the pandemic, this reference interaction was attempted to be replicated virtually. In many aspects, the interaction fidelity of the VR meetings was high. Action fidelity was relatively high, as people could use their hands for mid-air gestures, move around, and use head movements for turn-taking. Also, simulation fidelity was high in some regards, particularly scenario fidelity, because of the spatial relationships between people and the environment, which enables complex group dynamics, switching between conversations of subgroups, or guiding the group focus. The world's semantic relationships created a familiar environment for social gatherings and meetings, such as a presentation screen to gather around or attendees raising their hands to speak.

Regarding rendering fidelity, the visual realism was especially high on the platform Engage when using high-performance PC-VR systems. Additionally, the platforms' spatial audio renders the voices depending on the distance between interlocutors, allowing for simultaneous conversations of subgroups in one room and increasing rendering fidelity further. As meetings are mainly an audio-visual experience, we can consider display fidelity to be high. The VE extends to all sides and behind the users, allowing them to look around naturally. The fidelity of the audio and visual displays depends on the VR headset, as the attendees used devices from the budget Quest 1 to the sophisticated Valve Index. The participants' comments on experiential fidelity were highly diverse and sometimes depended on their previous VR experience and the hardware employed. The assessments ranged from a user describing the experience as "speaking on an old phone with multiple participants at the same time, wearing a rock tied to my head, and watching some meaningless cartoon simultaneously" to other participants describing the impression of sharing a common space, being in the office together, or using body language to communicate authentically.

Overall, the examined VR meetings had the potential for medium to high interaction fidelity when compared to the original in-person meeting before the pandemic. However, meeting in VR was no replacement for meeting in person but for meeting in video calls. Videoconferences were established at the beginning of the pandemic before planning the VR experiment. Since going back to in-person meetings was no option at the time, video conferences were the only safe alternative to VR meetings. Therefore, attempting to replicate the traditional meeting style from before the mandatory home office was not an ideal target reference interaction. Despite comparatively high interaction fidelity, attendees were frustrated and preferred returning to Zoom. We fell into the Apples-and-Oranges Trap described in **P1** (IntFi Model) by confusing what the appropriate reference interaction is and why our fidelity assessment is misleading.

For the meeting format examined in this case study, advantages such as the highly realistic spatial resemblance were not decisive for the success of the meetings. This was unnecessary for the primarily static one-to-many announcements or presentations with little audience interaction. Instead, the user experience was limited by technical issues, discomfort, motion sickness, and the restriction of secondary tasks. Ultimately, the guiding research question should not have been on the meeting fidelity compared to in-person meetings but on whether they are competitive with the ease and comfort of videoconferences.

This ambiguity of the two reference interactions was also reflected in the results. Participants also compared to both references. They shared appreciation for the feeling of "getting together" but missed the benefits of videoconferencing, e.g., multi-tasking or ease of use. Most importantly for the attendees, seeing each other's faces and recognizing emotions was impossible in VR. Facial expressions are not tracked by any of the VR systems used in the experiment. According to the Irrevocable-Loss Pattern from **P1** (IntFi Model), the resulting limitation in detection fidelity propagated throughout the loop without a possibility to compensate for it. With no information on the user's facial expression, it cannot be transferred into virtual actions, it cannot be considered in the simulation, it cannot be rendered or displayed, and eventually, it cannot be perceived by peers. At the end of the experiment, this was one of the main reasons for the team's decision to return to video conferences.

Further research with a clear separation of researchers and meeting attendees is needed. At the cost of the insights from an autoethnographic approach, the generalizability can be higher with controlled, balanced, and diverse samples across domains and prior experience. As the sample in this case study mainly consisted of HCI researchers with high technical literacy, the results are most likely not representative of the general population. A comparison in follow-up studies with groups of varying VR experience, media competence, demographics, work domains, or team sizes would be interesting. Also, other types of collaboration beyond static meetings could be the focus of further research to pave the way for immersive remote teamwork. Moreover, finding novel solutions and adapting VR technology with its unique possibilities beyond replicating in-person meetings could create unprecedented advantages for collaborating remotely.

6
Discussion & Outlook

The research presented in this dissertation demonstrates the fundamental impact of interaction fidelity in different domains on various parameters of VR applications, including user performance, satisfaction, perceived realism, sense of body ownership, engagement, agency, and cognitive load. This makes it an essential factor in interaction design and system development, as illustrated at the end of the previous chapters in response to the research questions. The faithful replication of the real world is decisive for the success of various use cases, including training, teleoperation, marketing, teaching, tourism, sports, digital conservation, and many others. Regarding **RQ1**, I outlined how providing the fidelity concept with structure and unambiguous definitions facilitates insightful analyses. The IntFi Model allows multi-dimensional comparability of fidelity across a vast range of VR interactions and use cases.

However, this thesis has also demonstrated that higher fidelity is not always better or more desirable. We have seen worse user performance with medium fidelity than with low fidelity in **P4** (Variable Grip), participants who could not perceive the adaptive trigger resistance in **P5** (Triggermuscle), increased task load for players in **P6** (Kicking), less courtesy towards the most realistic agent in **P8** (Agent Embodiment), and lower co-presence in **P9** (VR Meetings)—besides many positive effects. The level of fidelity must align with the users' and designers' goals. For example, comfort and ease of use were more important for the meeting format in **P9** (VR Meetings) than closely replicating in-person meetings' spatial characteristics and affordances. Furthermore, as we have seen in **P3** (in-VR Questionnaires) concerning **RQ2**, it is crucial to start with identifying a purposeful reference interaction that focuses on the goals of the simulation.

Several fidelity aspects were increased with the novel haptic interfaces for object manipulation introduced in **P4** (Variable Grip) and **P5** (Triggermuscle). Yet, the haptic impressions and object control were not as successful as hypothesized, which links to **RQ3**. Although action and haptic fidelity could be significantly increased with the force-feedback glove, the system falls short of users' expectations of maximum fidelity: The interface does not allow handling objects as if they were real. Interfaces that provide close to but not quite maximum interaction fidelity are in danger of being affected by what McMahan et al. (2016) called the uncanny valley of VR interactions. Furthermore, while action and output fidelity were higher with adaptive trigger resistance, some users did not detect the haptic stimuli or interpret them as weight. Weight perception with this mapping is unreliable when strong side effects such as vibrations occur. It is noteworthy how the participants' perceptions and experiences differed fundamentally, although all used the same imperfect system. This emphasizes the importance of considering users and target groups—not only system fidelity—when assessing interaction fidelity.

In response to **RQ4**, we observed strong effects of varying interaction fidelity in performing motor skills in VR sports, affecting both objective performance measures and subjective experiential measures in **P6** (Kicking) and **P7** (VRisbee). The increased rendering fidelity propagated through the IntFi Loop, resulting in higher experiential and action fidelity. Similarly, higher rendering fidelity of the conversational agents of **P8** (Agent Embodiment) elicited better user responses. However, the observed impact of varying interaction fidelity in videoconferencing and VR meetings was highly individual. On the one hand, the case study in **P9** (VR Meetings) found behaviors and experiences typical for the reference interactions of in-person meetings, such as using body language, forming queues, or using spatial audio when forming small conversation groups. On the other hand, for many participants, the system was not nearly close enough to in-person meetings to outweigh the disadvantages regarding inconvenience, multitasking, and effort.

Reproducing the real world is not always the most feasible solution given technological, economic, or practical limitations, such as computing power or tracking space constraints. Unrealistic solutions have been proven useful with countless effective interaction techniques (e.g., gaze-and-pinch selection), locomotion techniques (e.g., teleportation), perceptual manipulations (e.g., redirected grasping), virtual superpowers (e.g., suspend gravity), and sensory substitutions (e.g., vibrotactile feedback for collision with an object). These approaches often deliberately deviate from a direct replication in a particular fidelity component, requiring interesting compensation effects on other components. Users also adapt well to abstract mappings that do not aim to replicate anything but introduce new concepts. In many applications, not being limited by the constraints of reality is the crucial advantage of XR technology. Supernatural abilities and fictional worlds can inspire and serve as insightful reference interactions. Even without striving for high fidelity, the Interaction Fidelity Model can be a helpful tool to understand and reflect on such mechanisms.

I want to emphasize that the ethical implications of high-fidelity VR interactions are significant as the technology and its widespread adoption involve substantial risks. We must be cautious not to turn the beneficial immersion in virtual worlds into severe psychological or societal harm. A growing body of literature raises awareness of the risks of XR technology with increasing fidelity and provides recommendations (Madary and Metzinger, 2016; Slater et al., 2020). As with other digital media, escapism is a severe risk of using XR. The psychological post-exposure effects in the context of embodiment can be especially harmful regarding users' self-perception and body image. Another risk is the loss of connection to reality or confusion of realities, such as confusing the source of a memory (Mizuho et al., 2023) or XR systems being misused for malicious attacks (Tseng et al., 2022). Furthermore, already existing global inequalities and societal injustice could be further exacerbated. The already existing technological gap (Robinson et al., 2020) could comprise a "fidelity gap" in the future, as high-fidelity systems are expensive and access is unequally distributed. Lastly, the interaction design for XR needs to be accessible. The IntFi Model emphasizes the consideration of the user as part of interactions. Since we cannot understand every user's perception, experience, and behavior, a feasible alternative is understanding target user groups. Including a diverse sample in these considerations, not only average users, is essential to ensure accessibility and inclusion by design.

With this dissertation, my goal was to unravel the multi-layered construct of fidelity, bring structure to this seminal field of research, and better understand its influence on interactions with virtual objects and humans. I want to briefly reflect on how expedient my chosen approach was. The IntFi Model lays the foundation for analyzing and understanding all factors contributing to overall interaction fidelity. For many use cases, specialized models on single subcomponents are required for detailed assessments. As such, the Haptic Fidelity Framework facilitates a differentiated analysis of one modality of output fidelity. In the literature, there are several other specialized frameworks. Still, there is a need for more theoretical work to fill the gaps. The theoretical contributions of this dissertation answered some crucial questions and cleared uncertainties while paving the way for far more unresolved questions that call for further research. The proposed model is not the ultimate solution to all fidelity concerns. Instead, it will need to be adapted and evolve with future research and technological development. I hope the community will identify further patterns and establish extensive connections between study findings to eventually extend or revise the IntFi Model, similar to how I have taken up the pioneering work by McMahan (2011).

The empirical investigations shaped my understanding of VR interactions and informed the proposed theory. With this, the presented publications followed a typical cycle of scientific knowledge creation: collecting empirical data that informs a theoretical model, which helps interpret additional empirical data to refine the theory further. The mutual influence of user-centric empiricism, technological experimentation, methodological reflection, and theoretical considerations shaped this dissertation. My selection of research goals and investigated topics could have been more streamlined, focusing on either object manipulation, embodiment, social VR, or conversational interfaces instead of a combination of those. On the other hand, my curiosity guided me to these topics, and I thoroughly enjoyed doing such diverse research. Although the direction of my research was not always predictable along the way, I am content in hindsight about how it evolved as it led to interesting investigations.

Looking forward, I expect significant technological advances that will considerably increase interaction fidelity. The pace at which XR technology has evolved in recent years is remarkable. Until we reach the far-away point of maximum fidelity with the "ultimate display," as envisioned by Sutherland (1965), the IntFi Model and similar frameworks will remain relevant and helpful in navigating the evolving XR landscape. Recent developments in mixed reality with the Apple Vision Pro and the Meta Quest 3 suggest how omnipresent virtual elements and spatial computing may be in our everyday lives in the future.

Providing high-fidelity haptic feedback currently requires complex devices. Aside from interactions with force-feedback gloves and controllers, improvements in optical hand tracking and rich mixed-reality applications offer intriguing opportunities for similar interaction techniques. For convincing hand manipulation of objects, complex physics-based simulations are needed that infer object behavior from the position and properties of the individual fingers (Höll et al., 2018). As the contact points between fingers and surface change according to how the object moves within the hand, the adaption of the grasping pose must be determined dynamically, as previous work explored for freehand grasping (Dalia Blaga et al., 2021). For controlling the grip applied to an object's surface or perceiving its weight, the penetration depth could be used to indicate grip forces or mass in future studies. Similarly, more research is needed to understand better how to design intuitive and reliable interfaces that enable realistic object manipulation in terms of further physical properties.

Regarding research on embodiment and virtual humans, further improvements in rendering and simulation fidelity are essential for some use cases, such as social VR. Although the case study in **P9** (VR Meetings) clearly showed that XR technology was not sophisticated enough in 2020 to make it the primary choice for work-related meetings, I expect that technical advancements will facilitate a drastic increase in virtual encounters with family, friends, colleagues, and peers over the coming years. Also, the rapid progress in artificial intelligence technology enabling conversational interactions to be integrated into VEs has the potential to revolutionize assisting technologies, such as in **P8** (Agent Embodiment). This could lead to a greater demand for credible virtual agents and customized self-avatars that match our real-world appearance. Since achieving virtual humans with convincingly high fidelity in every aspect is expensive and complicated, further research on the contributing factors can help prioritize appropriately for different use cases on the journey to virtual humans with maximum fidelity.

As technology improves and develops the potential of generating increasingly realistic virtual and blended realities, research must accompany and inform the transition to a new interface era of spatial computing. **P1** (IntFi Model) outlines a substantial research agenda and further implications of fidelity research, including considerations of optimizing for realism, how fidelity relates to similar constructs, whether we can describe fidelity objectively, how it could be quantified and measured, and how it can be applied to other reference frames, such as fiction or MR. While existing frameworks and taxonomies examine various fidelity components, such as input fidelity (McMahan et al., 2016), haptic output fidelity as in **P2** (Haptic Fidelity), auditive output fidelity (Lindau et al., 2014), or experiential fidelity (Alexander et al., 2005; Lindeman and Beckhaus, 2009), specialized models are missing for other components. The particularly complex components of experiential and simulation fidelity especially require continued attention in research due to their complexity. On theoretical grounds, empirical findings can be generalized more meaningfully and guide our research agendas with foresight.

7
Conclusion

Understanding the fidelity of VR interactions is as challenging as it is helpful. Similar to indicators such as presence, user experience, or usability, fidelity is not the most suitable measure for every application or goal. However, reflecting on the reference interaction to be simulated and how closely it corresponds to the simulation can be insightful. Unraveling the factors determining fidelity allows researchers and developers to improve and prioritize system and interaction design.

This dissertation facilitates a better comprehension of fidelity and realism in VR by elaborating on theoretical foundations, methodological considerations, and empirical findings from user studies. All research was conducted in teams. Based on the insights from nine papers, this dissertation presents the following contributions as a response to the four research questions. The conceptual taxonomy of the Interaction Fidelity Model delineates and defines eight fidelity components, provides application guidelines, and proposes underlying patterns. The Haptic Fidelity Framework allows zooming into one of the subcomponents of this universal model for a more detailed evaluation of haptics. It enables a detailed quantitative assessment of the 14 factors determining the fidelity of haptic interfaces.

Furthermore, I presented the findings from a series of empirical studies. The research demonstrated the effectiveness and high usability of integrating questionnaires in VEs for seamless self-reporting in VR user studies. I elaborated on two projects investigating haptic interfaces that render physical properties when manipulating virtual objects. The evaluations of an interaction technique using a force-feedback glove and a hand-held controller with adaptive trigger resistance revealed intricacies of haptic perception and user experience for dexterous object manipulation. Moreover, I discussed the results of four studies on embodiment and virtual humans. This includes investigations of hand and foot visibility in VR sports and their impact on player performance, body ownership, and perceived control. Another study explored how users assess the embodiment of a voice assistant agent with varying levels of audio-visual realism. Lastly, I presented a case study providing genuine insights from testing social VR platforms for a team's weekly group meetings in comparison to videoconferences.

This thesis opens up abundant opportunities for further research. The proposed theoretical models can be used for meta-analyses to uncover additional common patterns of interaction fidelity and inform our understanding of the realism of VR interfaces. We refer to some specialized tools and theories on subcomponents of the IntFi Model. Still, there are numerous gaps where rigorous and detailed frameworks would be helpful for researchers, designers, and developers. Considering the swift pace of innovation in XR technology, we need continued research efforts on haptic interfaces, interaction techniques for realistic object manipulation, and the convincing embodiment of virtual humans for rich social interactions.

The empirical investigations have shown that the relationship between interaction fidelity and desired effects, such as user satisfaction, performance, task load, or perceived realism, is neither linear nor trivial. The intricate interplay of the fidelity components requires careful consideration, systematic analysis, and purposeful interaction design. This thesis provides helpful guidelines, detailed examples, analyses, empirical evidence, educational material, and assessment tools to make sense of the complex construct of interaction fidelity.

References

Hamza A. Al-Jundi and Emad Y. Tanbour. 2022. A framework for fidelity evaluation of immersive virtual reality systems. *Virtual Reality* (Jan. 2022). https://doi.org/10.1007/s10055-021-00618-y

Amy L Alexander, Tad Brunyé, Jason Sidman, Shawn A Weil, and others. 2005. From gaming to training: A review of studies on fidelity, immersion, presence, and buy-in and their effects on transfer in pc-based simulations and games. *DARWARS Training Impact Group* 5 (2005), 1–14.

Sean Andrist, Michael Gleicher, and Bilge Mutlu. 2017. Looking Coordinated: Bidirectional Gaze Mechanisms for Collaborative Interaction with Virtual Characters. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 2571–2582. https://doi.org/10.1145/3025453.3026033

Elisabeth André. 2011. Design and evaluation of embodied conversational agents for educational and advisory software. In *Gaming and Simulations: Concepts, Methodologies, Tools and Applications*. IGI Global, Hershey, Pennsylvania, USA, 668–686.

F. Argelaguet, L. Hoyet, M. Trico, and A. Lecuyer. 2016. The role of interaction in virtual embodiment: Effects of the virtual hand representation. In *2016 IEEE Virtual Reality (VR)*. 3–10. https://doi.org/10.1109/VR.2016.7504682

Anne-Sophie Augurelle, Allan M. Smith, Thierry Lejeune, and Jean-Louis Thonnard. 2003. Importance of Cutaneous Feedback in Maintaining a Secure Grip During Manipulation of Hand-Held Objects. *Journal of Neurophysiology* 89, 2 (Feb. 2003), 665–671. https://doi.org/10.1152/jn.00249.2002

Jeremy N. Bailenson, Kim Swinth, Crystal Hoyt, Susan Persky, Alex Dimov, and Jim Blascovich. 2005. The Independent and Interactive Effects of Embodied-Agent Appearance and Behavior on Self-Report, Cognitive, and Behavioral Markers of Copresence in Immersive Virtual Environments. *Presence: Teleoperators and Virtual Environments* 14, 4 (Aug. 2005), 379–393. https://doi.org/10.1162/105474605774785235

Miguel Barreda-Ángeles and Tilo Hartmann. 2021. Psychological benefits of using social virtual reality platforms during the covid-19 pandemic: The role of social and spatial presence. *Computers in Human Behavior* (Oct. 2021), 107047. https://doi.org/10.1016/j.chb.2021.107047

Steffi Beckhaus and Robert W. Lindeman. 2011. Experiential Fidelity: Leveraging the Mind to Improve the VR Experience. In *Virtual Realities: Dagstuhl Seminar 2008*, Guido Brunnett, Sabine Coquillart, and Greg Welch (Eds.). Springer, Vienna, 39–49. https://doi.org/10.1007/978-3-211-99178-7_3

Ayush Bhargava, Jeffrey W. Bertrand, Anand K. Gramopadhye, Kapil C. Madathil, and Sabarish V. Babu. 2018. Evaluating Multiple Levels of an Interaction Fidelity Continuum on Performance and Learning in Near-Field Training Simulations. *IEEE Transactions on Visualization and Computer Graphics* 24, 4 (April 2018), 1418–1427. https://doi.org/10.1109/TVCG.2018.2794639

Michael Bonfert, Robert Porzel, and Rainer Malaka. 2019. Get a Grip! Introducing Variable Grip for Controller-Based VR Systems. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 604–612. https://doi.org/10.1109/VR.2019.8797824

Michael Bonfert, Maximilian Spliethöver, Roman Arzaroli, Marvin Lange, Martin Hanci, and Robert Porzel. 2018. If You Ask Nicely: A Digital Assistant Rebuking Impolite Voice Commands. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction (ICMI '18)*. Association for Computing Machinery, New York, NY, USA, 95–102.

https://doi.org/10.1145/3242969.3242995

Matthew Botvinick and Jonathan Cohen. 1998. Rubber hands 'feel' touch that eyes see. *Nature* 391, 6669 (Feb. 1998), 756–756. https://doi.org/10.1038/35784

Doug A. Bowman, Ryan P. McMahan, and Eric D. Ragan. 2012. Questioning Naturalism in 3D User Interfaces. *Commun. ACM* 55, 9 (Sept. 2012), 78–88. https://doi.org/10.1145/2330667.2330687

Lila Bozgeyikli and Evren Bozgeyikli. 2019. Tangiball: Dynamic Embodied Tangible Interaction with a Ball in Virtual Reality. In *Companion Publication of the 2019 on Designing Interactive Systems Conference 2019 Companion (DIS '19 Companion)*. Association for Computing Machinery, New York, NY, USA, 135–140. https://doi.org/10.1145/3301019.3323904

Eric E. Brodie and Helen E. Ross. 1984. Sensorimotor Mechanisms in Weight Discrimination. *Perception & Psychophysics* 36, 5 (1984), 477–481. https://doi.org/10.3758/BF03207502

John Brooke. 1996. SUS: A Quick and Dirty Usability Scale. In *Usability Evaluation In Industry*, Patrick W. Jordan, B. Thomas, Ian Lyall McClelland, and Bernard Weerdmeester (Eds.). CRC Press, London; Bristol, PA, 189–194.

Karolis Butkus and Tautvydas Ceponis. 2019. Accuracy of throwing distance perception in virtual reality. In *IVUS*.

G. Cadoret and A. M. Smith. 1996. Friction, not texture, dictates grip forces used during object manipulation. *Journal of Neurophysiology* 75, 5 (May 1996), 1963–1969. https://doi.org/10.1152/jn.1996.75.5.1963

Ryan Canales, Aline Normoyle, Yu Sun, Yuting Ye, Massimiliano Di Luca, and Sophie Jörg. 2019. Virtual Grasping Feedback and Virtual Hand Ownership. In *ACM Symposium on Applied Perception 2019 (SAP '19)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3343036.3343132

Zekun Cao, Jason Jerald, and Regis Kopper. 2018. Visually-Induced Motion Sickness Reduction via Static and Dynamic Rest Frames. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 105–112. https://doi.org/10/gf39qh

Stephanie Carnell, Anna Miles, and Benjamin Lok. 2022. Evaluating Virtual Patient Interaction Fidelity With Advanced Communication Skills Learners. *Frontiers in Virtual Reality* 2 (2022). https://doi.org/10.3389/frvir.2021.801793

J. Cassell, T. Bickmore, M. Billinghurst, L. Campbell, K. Chang, H. Vilhjálmsson, and H. Yan. 1999. Embodiment in Conversational Interfaces: Rea. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '99)*. Association for Computing Machinery, New York, NY, USA, 520–527. https://doi.org/10.1145/302979.303150

Vuthea Chheang, Patrick Saalfeld, Tobias Huber, Florentine Huettl, Werner Kneist, Bernhard Preim, and Christian Hansen. 2019. Collaborative virtual reality for laparoscopic liver surgery training. In *2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. IEEE, 1–17.

Andreea Dalia Blaga, Maite Frutos-Pascual, Chris Creed, and Ian Williams. 2021. Freehand Grasping: An Analysis of Grasping for Docking Tasks in Virtual Reality. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. 749–758. https://doi.org/10.1109/VR50410.2021.00102

Xavier de Tinguy, Thomas Howard, Claudio Pacchierotti, Maud Marchal, and Anatole Lécuyer. 2020. WeATaViX: WEarable Actuated TAngibles for VIrtual Reality eXperiences. In *Haptics: Science, Technology, Applications (Lecture Notes in Computer Science)*, Ilana Nisky, Jess Hartcher-O'Brien, Michaël Wiertlewski, and Jeroen Smeets (Eds.). Springer International Publishing, Cham, 262–270. https://doi.org/10.1007/978-3-030-58147-3_29

Xavier de Tinguy, Claudio Pacchierotti, Mathieu Emily, Mathilde Chevalier, Aurelie Guignardat, Morgan Guillaudeux, Chloe Six, Anatole Lecuyer, and Maud Marchal. 2019. How Different Tangible and Virtual Objects Can Be While Still Feeling the Same?. In *2019 IEEE World Haptics Conference (WHC)*. IEEE, Tokyo, Japan, 580–585. https://doi.org/10.1109/WHC.2019.8816164

Alexander Diel, Sarah Weigelt, and Karl F. Macdorman. 2021. A Meta-analysis of the Uncanny Valley's Independent and Dependent Variables. *ACM Transactions on Human-Robot Interaction* 11, 1 (2021), 1:1–1:33. https://doi.org/10.1145/3470742

J. M. Elliott and K. J. Connolly. 1984. A Classification of Manipulative Hand Movements. *Developmental Medicine & Child Neurology* 26, 3 (June 1984), 283–296. https://doi.org/10.1111/j.1469-8749.1984.tb04445.x

Thomas Erickson, N. Sadat Shami, Wendy A. Kellogg, and David W. Levine. 2011. Synchronous interaction among hundreds: an evaluation of a conference in an avatar-based virtual environment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. Association for Computing Machinery, New York, NY, USA, 503–512. https://doi.org/10.1145/1978942.1979013

A. S. Fernandes and S. K. Feiner. 2016. Combating VR Sickness through Subtle Dynamic Field-of-View Modification. In *2016 IEEE Symposium on 3D User Interfaces (3DUI)*. 201–210. https://doi.org/10/gcmfn6

Kraig Finstad. 2010. The Usability Metric for User Experience. *Interacting with Computers* 22, 5 (Sept. 2010), 323–327. https://doi.org/10/ctr6r4

Daniel Freeman, Polly Haselton, Jason Freeman, Bernhard Spanlang, Sameer Kishore, Emily Albery, Megan Denne, Poppy Brown, Mel Slater, and Alecia Nickless. 2018. Automated Psychological Therapy Using Immersive Virtual Reality for Treatment of Fear of Heights: A Single-Blind, Parallel-Group, Randomised Controlled Trial. *The Lancet Psychiatry* 5, 8 (Aug. 2018), 625–632. https://doi.org/10/gd2h5n

Siegfried J. Gerathewohl. 1969. *Fidelity of Simulation and Transfer of Training: A Review of the Problem,*. Technical Report. Federal Aviation Administration Washington D C Office of Aviation Medicine. https://apps.dtic.mil/sti/citations/AD0706744

Adrien Girard, Maud Marchal, Florian Gosselin, Anthony Chabrier, François Louveau, and Anatole Lécuyer. 2016. HapTip: Displaying Haptic Shear Forces at the Fingertips for Multi-Finger Interaction in Virtual Environments. *Frontiers in ICT* 3 (April 2016). https://doi.org/10.3389/fict.2016.00006

Andrew M. Gordon, Hans Forssberg, R. S. Johansson, and Göran Westling. 1991. Visual Size Cues in the Programming of Manipulative Forces During Precision Grip. *Experimental Brain Research* 83, 3 (1991), 477–482.

J. Grubert, L. Witzani, E. Ofek, M. Pahud, M. Kranz, and P. O. Kristensson. 2018. Effects of Hand Representations for Typing in Virtual Reality. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 151–158. https://doi.org/10/gf39qd

Sandra G. Hart. 1986. *NASA Task Load Index (TLX). Volume 1.0; Paper and Pencil Package*. Technical Report. NASA.

David Hecht and Miriam Reiner. 2009. Sensory Dominance in Combinations of Audio, Visual and Haptic Stimuli. *Experimental Brain Research* 193, 2 (2009), 307–314. https://doi.org/10.1007/s00221-008-1626-z

Shirley Holt. 2023. Virtual reality, augmented reality and mixed reality: For astronaut mental health; and space tourism, education and outreach. *Acta Astronautica* 203 (Feb. 2023), 436–446. https://doi.org/10.1016/j.actaastro.2022.12.016

Kasper Hornbæk and Antti Oulasvirta. 2017. What Is Interaction?. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 5040–5052. https://doi.org/10.1145/3025453.3025765

M. Hubbard and S. A. Hummel. 2000. Simulation of Frisbee Flight. In *Proceedings of the 5th Conference on Mathematics and Computers in Sport*, G. Cohen (Ed.). Sydney, Australia, 14–16.

Sarah Ann Hummel. 2003. *Frisbee Flight Simulation and Throw Biomechanics*. Master's thesis. University of California, Davis.

Markus Höll, Markus Oberweger, Clemens Arth, and Vincent Lepetit. 2018. Efficient Physics-Based Implementation for Realistic Hand-Object Interaction in Virtual Reality. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 175–182. https://doi.org/10.1109/VR.2018.8448284

Wijnand A IJsselsteijn, Yvonne A. W de Kort, and Antal Haans. 2006. Is This My Hand I See Before Me? The Rubber Hand Illusion in Reality, Virtual Reality, and Mixed Reality. *Presence: Teleoperators and Virtual Environments* 15, 4 (Aug. 2006), 455–464. https://doi.org/10.1162/pres.15.4.455

Katherine Isbister and Patrick Doyle. 2002. Design and evaluation of embodied conversational agents: A proposed taxonomy. In *The first international joint conference on autonomous agents & multi-agent systems*. ACM, NY, USA.

Robert J.K. Jacob, Audrey Girouard, Leanne M. Hirshfield, Michael S. Horn, Orit Shaer, Erin Treacy Solovey, and Jamie Zigelbaum. 2008. Reality-based interaction: a framework for post-WIMP interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*. Association for Computing Machinery, New York, NY, USA, 201–210. https://doi.org/10.1145/1357054.1357089

Jason Jerald. 2016. *The VR Book: Human-Centered Design for Virtual Reality* (first edition ed.). Number 8 in ACM Books. acm, Association for Computing Machinery, New York.

Lynette A. Jones and Hong Z. Tan. 2013. Application of Psychophysical Techniques to Haptic Research. *IEEE Transactions on Haptics* 6, 3 (July 2013), 268–284. https://doi.org/10.1109/TOH.2012.74

HyeongYeop Kang, Geonsun Lee, Seongsu Kwon, Ohung Kwon, Seongpil Kim, and JungHyun Han. 2018. Flotation Simulation in a Cable-Driven Virtual Environment – A Study with Parasailing. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 632:1–632:11. https://doi.org/10.1145/3173574.3174206

Behrang Keshavarz and Heiko Hecht. 2011. Validating an Efficient Method to Quantify Motion Sickness. *Human Factors* 53, 4 (Aug. 2011), 415–426. https://doi.org/10/dz78cg

Rabia Khan and Antonella De Angeli. 2009. The attractiveness stereotype in the evaluation of embodied conversational agents. In *IFIP Conference on Human-Computer Interaction*. Springer, Heidelberg, Germany, 85–97.

Konstantina Kilteni, Raphaela Groten, and Mel Slater. 2012. The Sense of Embodiment in Virtual Reality. *Presence: Teleoperators and Virtual Environments* 21, 4 (Nov. 2012), 373–387. https://doi.org/10.1162/PRES_a_00124

Erin Kim and Oliver Schneider. 2020. Defining Haptic Experience: Foundations for Understanding, Communicating, and Evaluating HX. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–13. https://doi.org/10.1145/3313831.3376280

Myung Jin Kim, Neung Ryu, Wooje Chang, Michel Pahud, Mike Sinclair, and Andrea Bianchi. 2022. SpinOcchio: Understanding Haptic-Visual Congruency of Skin-Slip in VR with a Dynamic Grip Controller. In *CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3491102.3517724

Bret Kinsella and Ava Mutchler. 2019. U.S. Smart Speaker Consumer Adoption Report 2019. https://voicebot.ai/smart-speaker-consumer-adoption-report-2019/

Kathrin Kirchner and Britta Nordin Forsberg. 2021. A Conference Goes Virtual: Lessons from Creating a Social Event in the Virtual Reality. In *Innovations for Community Services*, Udo R. Krieger, Gerald Eichler, Christian Erfurth, and Günter Fahrnberger (Eds.). Vol. 1404. Springer International Publishing, Cham, 123–134. https://doi.org/10.1007/978-3-030-75004-6_9

Jarrod Knibbe, Jonas Schjerlund, Mathias Petraeus, and Kasper Hornb\textbackslash{ae}k. 2018. The Dream Is Collapsing: The Experience of Exiting VR. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 483:1–483:13. https://doi.org/10.1145/3173574.3174057

Felix Kosmalla, André Zenner, Corinna Tasch, Florian Daiber, and Antonio Krüger. 2020. The Importance of Virtual Hands and Feet for Virtual Reality Climbing. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3334480.3383067

Saadi Lahlou, Roy Pea, Maxi Heitmayer, Martha G. Russell, Robin Schimmelpfennig, Paulius Yamin, Marina Everri, Antoine Cordelois, and Adelaide P. Dawes. 2021. Are we 'Beyond being there' yet?: Towards better interweaving epistemic and social aspects of virtual reality conferencing. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–6. https://doi.org/10.1145/3411763.3451579

Michael Lankes, Regina Bernhaupt, and Manfred Tscheligi. 2007. An Experimental Setting to Measure Contextual Perception of Embodied Conversational Agents. In *Proceedings of the International Conference on Advances in Computer Entertainment Technology (ACE '07)*. Association for Computing Machinery, New York, NY, USA, 56–59. https://doi.org/10.1145/1255047.1255058

Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. 2017. *Research Methods in Human-Computer Interaction* (2 ed.). Morgan Kaufmann, Cambridge, MA.

Alexander Lindau, Vera Erbes, Steffen Lepa, Hans-Joachim Maempel, Fabian Brinkman, and Stefan Weinzierl. 2014. A Spatial Audio Quality Inventory (SAQI). *Acta Acustica united with Acustica* 100, 5 (Sept. 2014), 984–994. https://doi.org/10.3813/AAA.918778

Robert W. Lindeman and Steffi Beckhaus. 2009. Crafting memorable VR experiences using experiential fidelity. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology (VRST '09)*. Association for Computing Machinery, New York, NY, USA, 187–190. https://doi.org/10.1145/1643928.1643970

Jack M. Loomis and Susan J. Lederman. 1986. Tactual Perception. In *Handbook of Perception and Human Performances*. Vol. 2. John Wiley & Sons, 1–41.

Ke Ma and Bernhard Hommel. 2015. The role of agency for perceived ownership in the virtual hand illusion. *Consciousness and Cognition* 36 (2015), 277–288. https://doi.org/10.1016/j.concog.2015.07.008

Karl F MacDorman, Robert D Green, Chin-Chang Ho, and Clinton T Koch. 2009. Too real for comfort? Uncanny responses to computer generated faces. *Computers in human behavior* 25, 3 (2009), 695–710.

Michael Madary and Thomas K. Metzinger. 2016. Real Virtuality: A Code of Ethical Conduct. Recommendations for Good Scientific Practice and the Consumers of VR-Technology. *Frontiers in Robotics and AI* 3 (2016). https://doi.org/10/gc6znw

K. Mania, D. Wooldridge, M. Coxon, and A. Robinson. 2006. The effect of visual and interaction fidelity on spatial cognition in immersive virtual environments. *IEEE Transactions on Visualization and Computer Graphics* 12, 3 (2006), 396–404. https://doi.org/10.1109/TVCG.2006.55

Douglas I. McCloskey. 1974. Muscular and Cutaneous Mechanisms in the Estimation of the Weights of Grasped Objects. *Neuropsychologia* 12, 4 (1974), 513–520. https://doi.org/10.1016/0028-3932(74)90081-5

Ryan Patrick McMahan. 2011. *Exploring the Effects of Higher-Fidelity Display and Interaction for Virtual Reality Games*. Ph. D. Dissertation. Virginia Tech. https://vtechworks.lib.vt.edu/handle/10919/30123

Ryan P. McMahan, Chengyuan Lai, and Swaroop K. Pal. 2016. Interaction Fidelity: The Uncanny Valley of Virtual Reality Interactions. In *Virtual, Augmented and Mixed Reality (Lecture Notes in Computer Science)*, Stephanie Lackey and Randall Shumaker (Eds.). Springer International Publishing, Cham, 59–70.

Joshua McVeigh-Schultz and Katherine Isbister. 2021. The Case for "Weird Social" in VR/XR: A Vision of Social Superpowers Beyond Meatspace. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3411763.3450377

Merriam-Webster. 2024. Definition of Fidelity. https://www.merriam-webster.com/dictionary/fidelity

Takato Mizuho, Takuji Narumi, and Hideaki Kuzuoka. 2023. Effects of the Visual Fidelity of Virtual Environments on Presence, Context-dependent Forgetting, and Source-monitoring Error. *IEEE Transactions on Visualization and Computer Graphics* 29, 5 (May 2023), 2607–2614. https://doi.org/10.1109/TVCG.2023.3247063

Masahiro Mori. 1970. The uncanny valley. *Energy* 7, 4 (1970), 33–35.

Seyyed Ahmad Mousavi, Mehdi Shahbazi, Elahe Arabameri, and Elham Shirzad. 2018. The Effect of Virtual Reality Training on Learning and Kinematics Characteristics of Dart Throwing. *International Journal of School Health* (2018).

Fares Moustafa and Anthony Steed. 2018. A longitudinal study of small group interaction in social virtual reality. In

Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology (VRST '18). Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3281505.3281527

Clifford Nass, Jonathan Steuer, and Ellen R. Tauber. 1994. Computers Are Social Actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '94)*. Association for Computing Machinery, New York, NY, USA, 72–78. https://doi.org/10.1145/191666.191703

Donald A. Norman and Stephen W. Draper. 1986. *User centered system design: new perspectives on human-computer interaction*. L. Erlbaum, Hillsdale (N. J.) London.

Sebastian Oberdörfer, David Heidrich, and Marc Erich Latoschik. 2019. Usability of Gamified Knowledge Learning in VR and Desktop-3D. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, Glasgow, Scotland Uk, 175:1–175:13. https://doi.org/10/gf3k86

Marianna Obrist, Sue Ann Seah, and Sriram Subramanian. 2013. Talking about Tactile Experiences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1659–1668. https://doi.org/10.1145/2470654.2466220

Randy Pausch, Jon Snoddy, Robert Taylor, Scott Watson, and Eric Haseltine. 1996. Disney's Aladdin: first steps toward storytelling in virtual reality. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques (SIGGRAPH '96)*. Association for Computing Machinery, New York, NY, USA, 193–203. https://doi.org/10.1145/237170.237257

Astrid Krogh Pedersen, Halfdan Isaksen, Mikkel Bjerregaard Riedel, Morten Jorgensen, Razvan Paisa, and Niels Christian Nilsson. 2023. Get a Variable Grip: A Comparison of Three Gripping Techniques for Controller-Based Virtual Reality. In *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 98–102. https://doi.org/10.1109/VRW58643.2023.00026

Vrushank Phadnis, Kristin Moore, and Mar Gonzalez-Franco. 2023. The Work Avatar Face-Off: Knowledge Worker Preferences for Realism in Meetings. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 960–969. https://doi.org/10.1109/ISMAR59233.2023.00112

Sandra Poeschl. 2017. Virtual reality training for public speaking—a QUEST-VR framework validation. *Frontiers in ICT* 4 (2017), 13.

Michael I. Posner, Mary J. Nissen, and Raymond M. Klein. 1976. Visual Dominance: An Information-Processing Account of its Origins and Significance. *Psychological Review* 83, 2 (1976), 157–171. https://doi.org/10.1037/0033-295X.83.2.157

Susanne Putze, Dmitry Alexandrovsky, Felix Putze, Sebastian Höffner, Jan D. Smeddinck, and Rainer Malaka. 2020. Breaking The Experience: Effects of Questionnaires in VR User Studies. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Honolulu, HI, USA. https://doi.org/10.1145/3313831.3376144

Maria Pyasik, Gaetano Tieri, and Lorenzo Pia. 2020. Visual appearance of the virtual hand affects embodiment in the virtual hand illusion. *Scientific Reports* 10, 1 (March 2020), 5412. https://doi.org/10.1038/s41598-020-62394-0

John R. Raser. 1969. *Simulation and Society: An Exploration of Scientific Gaming*. Allyn and Bacon.

Philipp A. Rauschnabel, Reto Felix, Chris Hinsch, Hamza Shahab, and Florian Alt. 2022. What is XR? Towards a Framework for Augmented and Virtual Reality. *Computers in Human Behavior* 133 (Aug. 2022), 107289. https://doi.org/10.1016/j.chb.2022.107289

Martin Riemer, Jörg Trojan, Marta Beauchamp, and Xaver Fuchs. 2019. The rubber hand universe: On the impact of methodological differences in the rubber hand illusion. *Neuroscience & Biobehavioral Reviews* 104 (Sept. 2019), 268–280. https://doi.org/10.1016/j.neubiorev.2019.07.008

L. Robinson, J. Schulz, G. Blank, M. Ragnedda, H. Ono, B. Hogan, G. S. Mesch, S. R. Cotten, S. B. Kretchmer, T. M. Hale, T. Drabowicz, P. Yan, B. Wellman, M.-G. Harper, A. Quan-Haase, H. S. Dunn, A. A. Casilli, P. Tubaro, R. Carvath, W. Chen, J. B. Wiest, M. Dodel, M. J. Stern, C. Ball, K.-T. Huang, and A. Khilnani. 2020. Digital inequalities 2.0: legacy inequalities in the information age. *First Monday* 25, 7 (2020). https://ora.ox.ac.uk/objects/uuid:

ac36d8d5-cd59-4871-ab30-7925b6714243

Katja Rogers, Sukran Karaosmanoglu, Maximilian Altmeyer, Ally Suarez, and Lennart E. Nacke. 2022. Much Realistic, Such Wow! A Systematic Literature Review of Realism in Digital Games. In *CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–21. https://doi.org/10.1145/3491102.3501875

A. F. Rovers and H.A. Essen, van. 2006. Guidelines for haptic interpersonal communication applications : an exploration of foot interaction styles. *Virtual Reality* 9, 2-3 (2006), 177–191. https://doi.org/10.1007/s10055-005-0016-0

S. V. Salazar, C. Pacchierotti, X. de Tinguy, A. Maciel, and M. Marchal. 2020. Altering the Stiffness, Friction, and Shape Perception of Tangible Objects in Virtual Reality Using Wearable Haptics. *IEEE Transactions on Haptics* 13, 1 (Jan. 2020), 167–174. https://doi.org/10.1109/TOH.2020.2967389

Evren Samur. 2012. Performance Evaluation Based on Physical Measurements. In *Performance Metrics for Haptic Interfaces*. Springer London, London, 43–65. https://doi.org/10.1007/978-1-4471-4225-6_4

Daniel Sarkady, Larissa Neuburger, and Roman Egger. 2021. Virtual reality as a travel substitution tool during COVID-19. In *Information and communication technologies in tourism 2021*. Springer, 452–463.

Anthony Scavarelli, Ali Arya, and Robert J. Teather. 2021. Virtual reality and augmented reality in social learning spaces: a literature review. *Virtual Reality* 25, 1 (March 2021), 257–277. https://doi.org/10.1007/s10055-020-00444-8

Ulrike Schultze. 2010. Embodiment and Presence in Virtual Worlds: A Review. *Journal of Information Technology* 25, 4 (Dec. 2010), 434–449. https://doi.org/10.1057/jit.2009.25

Peter Schulz, Dmitry Alexandrovsky, Felix Putze, Rainer Malaka, and Johannes Schöning. 2019. The Role of Physical Props in VR Climbing Environments. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300413

Valentin Schwind, Pascal Knierim, Nico Haas, and Niels Henze. 2019. Using Presence Questionnaires in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–12. https://doi.org/10.1145/3290605.3300590

Valentin Schwind, Sven Mayer, Alexandre Comeau-Vermeersch, Robin Schweigert, and Niels Henze. 2018. Up to the Finger Tip: The Effect of Avatars on Mid-Air Pointing Accuracy in Virtual Reality. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '18)*. ACM, Melbourne, VIC, Australia, 477–488. https://doi.org/10/gf3ngj

Stefania Serafin, Ali Adjorlu, Niels Nilsson, Lui Thomsen, and Rolf Nordahl. 2017. Considerations on the use of virtual and augmented reality technologies in music education. In *2017 IEEE virtual reality workshop on K-12 embodied learning through virtual & augmented reality (KELVAR)*. IEEE, 1–4.

Richard Skarbez, Frederick P. Brooks, Jr., and Mary C. Whitton. 2017. A Survey of Presence and Related Concepts. *Comput. Surveys* 50, 6 (Nov. 2017), 96:1–96:39. https://doi.org/10.1145/3134301

Mel Slater, Cristina Gonzalez-Liencres, Patrick Haggard, Charlotte Vinkers, Rebecca Gregory-Clarke, Steve Jelley, Zillah Watson, Graham Breen, Raz Schwarz, William Steptoe, Dalila Szostak, Shivashankar Halan, Deborah Fox, and Jeremy Silver. 2020. The Ethics of Realism in Virtual and Augmented Reality. *Frontiers in Virtual Reality* 1 (2020). https://doi.org/10.3389/frvir.2020.00001

Mel Slater and Maria V. Sanchez-Vives. 2016. Enhancing Our Lives with Immersive Virtual Reality. *Frontiers in Robotics and AI* 3 (2016). https://doi.org/10.3389/frobt.2016.00074

Carolin Stellmacher, André Zenner, Oscar Javier Ariza Nunez, Ernst Kruijff, and Johannes Schöning. 2023. Continuous VR Weight Illusion by Combining Adaptive Trigger Resistance and Control-Display Ratio Manipulation. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. 243–253. https://doi.org/10.1109/VR55154.2023.00040

Thomas A Stoffregen, Benoit G Bardy, LJ Smart, and RJ Pagulayan. 2003. On the nature and evaluation of fidelity in virtual environments. *Virtual and adaptive environments: Applications, implications, and human performance issues* (2003), 111–128.

Ivan E. Sutherland. 1965. The Ultimate Display. In *Proceedings of the IFIP Congress*. 506–508.

Judith Tirp, Christina Steingröver, Nick Wattie, Joe Baker, and Jörg Schorer. 2015. Virtual realities as optimal learning environments in sport – A transfer study of virtual and real dart throwing. *Psychological Test and Assessment Modeling* 57 (Jan. 2015), 57–69.

Wen-Jie Tseng, Elise Bonnail, Mark McGill, Mohamed Khamis, Eric Lecolinet, Samuel Huron, and Jan Gugenheimer. 2022. The Dark Side of Perceptual Manipulations in Virtual Reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3491102.3517728

Vonne Van Polanen and Marco Davare. 2015. Sensorimotor Memory Biases Weight Perception During Object Lifting. *Frontiers in Human Neuroscience* 9 (2015), 1–11. https://doi.org/10.3389/fnhum.2015.00700

Isaac Wang, Jesse Smith, and Jaime Ruiz. 2019. Exploring Virtual Agents for Augmented Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3290605.3300511

G. Westling and R. S. Johansson. 1984. Factors influencing the force control during precision grip. *Experimental Brain Research* 53, 2 (Jan. 1984), 277–284. https://doi.org/10.1007/BF00238156

Eric Whitmire, Hrvoje Benko, Christian Holz, Eyal Ofek, and Mike Sinclair. 2018. Haptic Revolver: Touch, Shear, Texture, and Shape Rendering on a Reconfigurable Virtual Reality Controller. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. ACM Press, Montreal QC, Canada, 1–12. https://doi.org/10.1145/3173574.3173660

C. Wienrich, K. Schindler, N. Döllinqer, S. Kock, and O. Traupe. 2018. Social Presence and Cooperation in Large-Scale Multi-User Virtual Reality - The Relevance of Social Interdependence for Location-Based Environments. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 207–214. https://doi.org/10/gf39qb

Julie Williamson, Jie Li, Vinoba Vinayagamoorthy, David A. Shamma, and Pablo Cesar. 2021. Proxemics and Social Interactions in an Instrumented Virtual Reality Workshop. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3411764.3445729

Bob G. Witmer and Michael J. Singer. 1998. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoper. Virtual Environ.* 7, 3 (June 1998), 225–240. https://doi.org/10.1162/105474698565686

Ye Yuan and Anthony Steed. 2010. Is the rubber hand illusion induced by immersive virtual reality?. In *2010 IEEE Virtual Reality Conference (VR)*. 95–102. https://doi.org/10.1109/VR.2010.5444807

Shumin Zhai. 1996. *Human performance in six degree of freedom input control*. PhD Thesis. University of Toronto, Ottawa.

Katja Zibrek and Rachel McDonnell. 2019. Social presence and place illusion are affected by photorealism in embodied VR. In *Motion, Interaction and Games (MIG '19)*. Association for Computing Machinery, New York, NY, USA, 1–7. https://doi.org/10.1145/3359566.3360064

Tim Zindulka, Myroslav Bachynskyi, and Jörg Müller. 2020. Performance and Experience of Throwing in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3313831.3376639

Part II

Publications

8

Original Publications

This chapter contains all nine papers of this dissertation in their original form, with content and formatting as published. All research has been conducted and reported in teams. On the cover page of each paper, I list all authors, summarize the content, and state my personal contributions to the work along the *CRediT* taxonomy. At the end of the chapter is a list of all my publications to date, including work outside the scope of this dissertation.

Contribution Taxonomy *CRediT*

These are the 14 roles typically played by contributors to research outputs according to the *CRediT* taxonomy.[1] The taxonomy has been refined by Consortia Advancing Standards in Research Administration (CASRAI) and National Information Standards Organization (NISO). I refer to these roles when outlining my contributions.

Conceptualization: Ideas; formulation or evolution of overarching research goals and aims.

Data curation: Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later re-use.

Formal analysis: Application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize study data.

Funding acquisition: Acquisition of the financial support for the project leading to this publication.

Investigation: Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection.

Methodology: Development or design of methodology; creation of models.

Project administration: Management and coordination responsibility for the research activity planning and execution.

Resources: Provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools.

Software: Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components.

Supervision: Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team.

Validation: Verification, whether as a part of the activity or separate, of the overall replication/reproducibility of results/experiments and other research outputs.

Visualization: Preparation, creation and/or presentation of the published work, specifically visualization/data presentation.

Writing – original draft: Preparation, creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation).

Writing – review & editing: Preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary or revision – including pre- or post-publication stages.

[1] https://credit.niso.org/contributor-roles-defined

Publication P1

The Interaction Fidelity Model: A Taxonomy to Distinguish the Aspects of Fidelity in Virtual Reality

Michael Bonfert, Thomas Muender, Ryan P. McMahan, Frank Steinicke, Doug Bowman, Rainer Malaka, and Tanja Döring

This publication introduces a model to determine how faithful VR interactions are to reference interactions. Based on the HCI Loop, the IntFi Model systematically covers all stages of VR interactions. It establishes a clear structure and precise definitions of eight distinct components. The conceptual model was reviewed through interviews with fourteen VR experts. We provide guidelines, diverse examples, and educational material to apply the IntFi Model universally to any VR experience. We identify common patterns and propose foundational research opportunities.

My contribution: I contributed the research idea and the majority of the theoretical considerations (*conceptualization*), the literature review, and most of the writing of the manuscript (*writing – draft, review, & editing*). I designed and conducted the validation method (*data curation* and *methodology*), conducted the expert interviews (*investigation*), and analyzed the qualitative data (*formal analysis*). I created most of the figures and the slide deck, and contributed to the other figures and the poster (*visualization*). I coordinated the project (*project administration*).

Preprint on *arXiv*, arXiv:2402.16665v1 [cs.HC], 2024

The Interaction Fidelity Model: A Taxonomy to Distinguish the Aspects of Fidelity in Virtual Reality

Michael Bonfert
bonfert@uni-bremen.de
University of Bremen
Germany

Thomas Muender
thom@uni-bremen.de
University of Bremen
Germany

Ryan P. McMahan
rpm@ucf.edu
University of Central Florida
USA

Frank Steinicke
frank.steinicke@uni-hamburg.de
Universität Hamburg
Germany

Doug Bowman
dbowman@vt.edu
Virginia Tech
USA

Rainer Malaka
malaka@uni-bremen.de
University of Bremen
Germany

Tanja Döring
tanja.doering@uni-bremen.de
University of Bremen
Germany

ABSTRACT

Fidelity describes how closely a replication resembles the original. It can be helpful to analyze how faithful interactions in virtual reality (VR) are to a reference interaction. In prior research, fidelity has been restricted to the simulation of reality—also called realism. Our definition includes other reference interactions, such as superpowers or fiction. Interaction fidelity is a multilayered concept. Unfortunately, different aspects of fidelity have either not been distinguished in scientific discourse or referred to with inconsistent terminology. Therefore, we present the Interaction Fidelity Model (IntFi Model). Based on the human-computer interaction loop, it systematically covers all stages of VR interactions. The conceptual model establishes a clear structure and precise definitions of eight distinct components. It was reviewed through interviews with fourteen VR experts. We provide guidelines, diverse examples, and educational material to universally apply the IntFi Model to any VR experience. We identify common patterns and propose foundational research opportunities.

CCS CONCEPTS

• **Human-centered computing → HCI theory, concepts and models**; **Virtual reality**; • **Computing methodologies** → *Virtual reality*.

KEYWORDS

VR, fidelity, realism, theory, framework, HCI, input, simulation, output

This is a preprint that has not yet been published in a peer-reviewed journal. The manuscript is currently under review. Current version of figures, definitions, and supplemental material: v2.1

1 INTRODUCTION

Realism in virtual reality (VR) is pursued intensely in research and development [16, 31, 40, 78, 84, 109, 112]. While the concept of *realism*—how closely a simulation resembles reality—may initially seem straightforward, it is a complex, multi-faceted construct. We quickly assess something as realistic or unrealistic, be it a painting, the behavior of a movie character, a synthetic voice, or a virtual world. However, this intuitive judgment is insufficient for a comprehensive understanding of and reasoning why something is perceived as more or less realistic, especially in such a complex domain as VR, where countless factors might influence the outcome. To purposefully design virtual experiences that are convincingly realistic, it is essential to untangle the different aspects that impact the overall realism.

VR technology can create immersive experiences of being in and interacting with simulated realities. By interacting with the VR system, a user can perceive and affect the virtual environment (VE), while the system can sense and react to user input. As in the real world, users and their environments can mutually influence each other. For many VR applications, realism is a decisive quality metric. The true-to-life resemblance is essential for skill training (e.g., surgery [20]), learning abilities (e.g., sports climbing [91], vocational education like public speaking [75], music [92]), entertainment (e.g., traveling the world [89]), therapy (e.g., fear of heights [27]), or use cases that would be expensive or impossible without VR (e.g., visiting Mars [36]). In these scenarios, the success of the simulation depends on how closely the equivalent from reality can be reproduced. Even in fictional scenarios, certain aspects of the interaction might need to be grounded in reality (e.g., Euclidean geometry, spatial audio, swarm behavior, gravity, or the color space perceptible by humans). However, the concept of realism is limited to matching the real world and, therefore, cannot be applied to VR use cases simulating aspects impossible in reality.

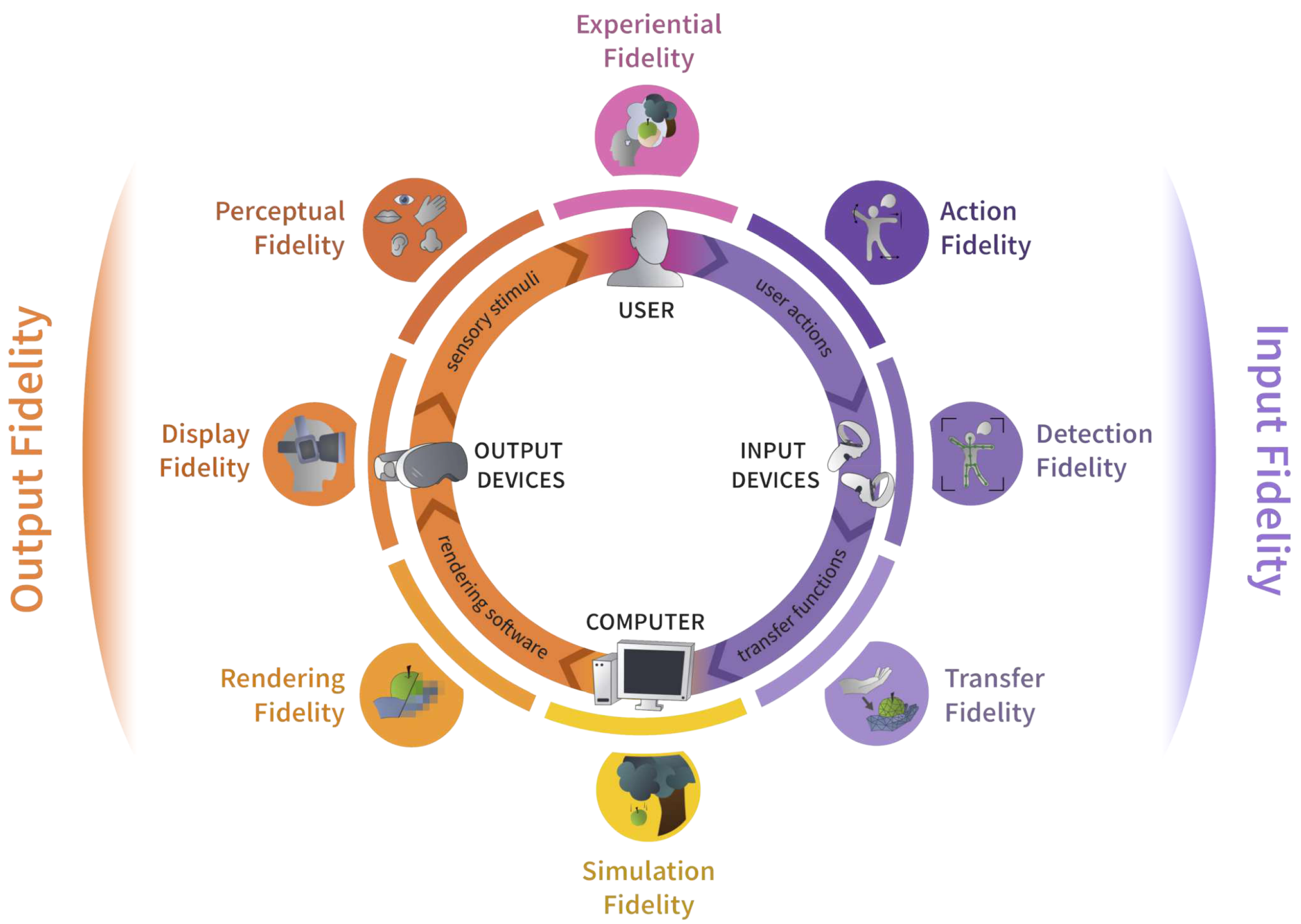


Figure 1: The Interaction Fidelity Model differentiates between eight fidelity components that affect any VR interaction. The model is based on the Human–Computer Interaction Loop [71, 81] and extends a previous framework by McMahan et al. [59].

1.1 The Concept of Fidelity

More generally speaking, the degree of how accurately an original is reproduced is called *fidelity* [63]. Every VR simulation recreates some reference, be it a real-life situation, a training scenario, a fictional world, or a designer's imagination. From the Latin term *fidēlis* for "faithful," fidelity describes how faithful something is to its original. When simulating reality, this degree of correspondence is called realism. Thus, realism is a specific form of fidelity. Therefore, a VR application can have low realism yet high fidelity to a reference frame other than the real world. For example, the swords in *Beat Saber* have high fidelity to lightsabers from *Star Wars* but are unrealistic. Comparing simulations more generally to reference systems than only to the real world has been suggested by Raser [79] in 1969, and we adopt this notion for more universal applicability. Figure 2 compares reference interactions and their simulated VR interactions in two examples. On the top, a real-life activity is reproduced virtually: picking an apple from a tree. This has been investigated with haptic interfaces of different realism [23, 43]. On the bottom, the fantasy concept of flying with a magic carpet is adapted as a locomotion technique [73]. This metaphor affords a large design space of possible realizations with varying interaction fidelity in VR [62]. The terms fidelity, realism, and naturalness are often used universally in scientific literature without detailing which aspect is referred to. If a specific aspect of fidelity is mentioned, a clear definition is often missing [84]. As a result, the terms and definitions within the VR literature have been inconsistent and contradictory, as illustrated in Section 2.

While the community's research efforts have led to useful definitions, models, and frameworks (see, e.g., Table 2), these mainly focus on dedicated aspects of fidelity, differ in their use of terms, and therefore do not provide a comprehensive understanding. This makes it harder to establish links between individual discoveries, generalize the results, and synthesize fundamental principles. Thus, this research aims to provide an umbrella framework that conflates existing findings on different aspects of fidelity into one comprehensive and consistent model.

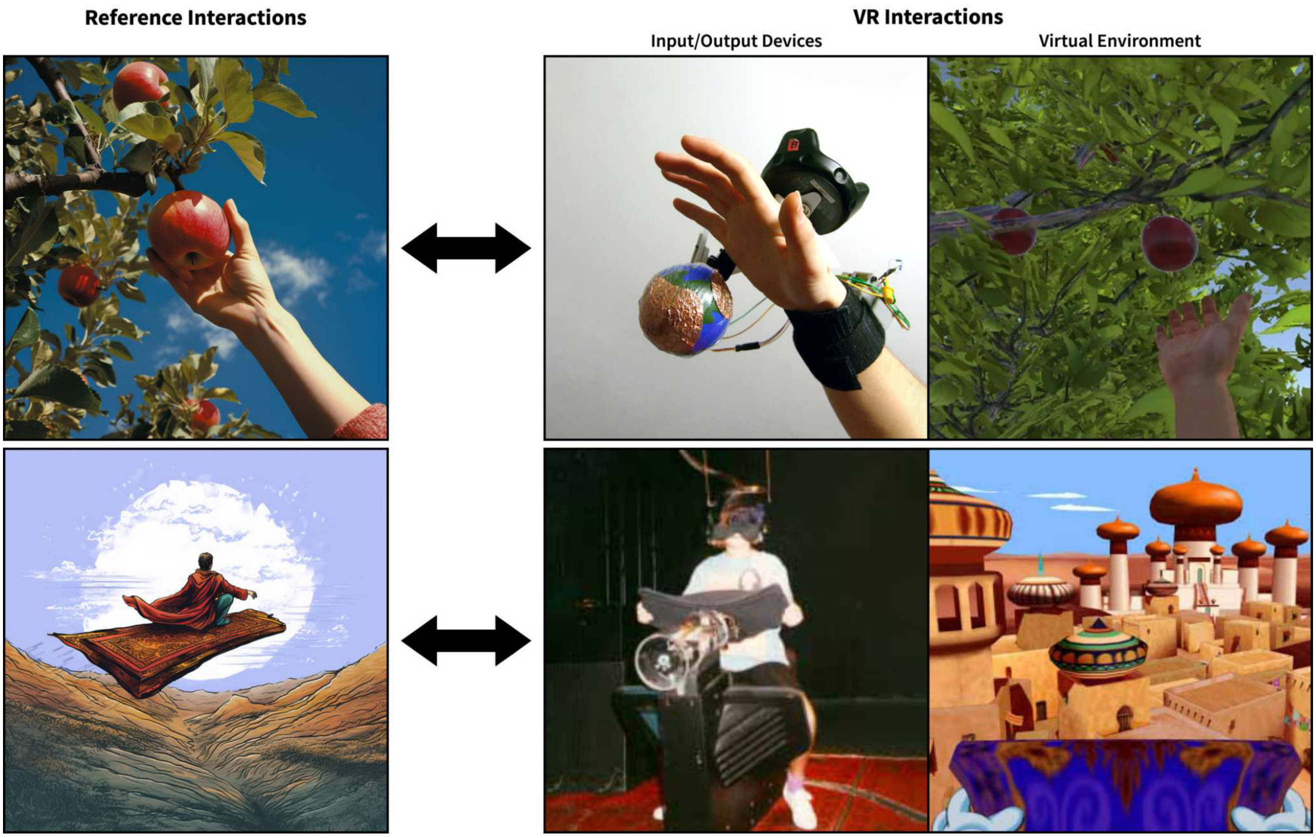


Figure 2: Two examples of reference interactions that are virtually replicated in VR simulations. *(top)* Here, an interaction from the real world is reproduced. Picking an apple from a tree is a complex activity to be simulated virtually as the user's actions, the provided feedback, and the physical simulation are closely coupled. This is a recent implementation with a haptic device by de Tinguy et al. [23] which dynamically moves the sphere proxy into the user's hand when grasping. *(bottom)* Here, a fictional interaction is reproduced. The mythological *magic carpet* from Middle Eastern literature can be realized as a VR locomotion technique in many different ways [62]. This is an early implementation by Pausch et al. [73]. *Copyright of the top-right images by [23] and of the bottom-right images by [73] (modified by the authors). The images for the reference interactions have been generated with Midjourney.*

1.2 Introducing the Interaction Fidelity Model

Therefore, we present the Interaction Fidelity Model (IntFi Model). This conceptual model distinguishes the different aspects of fidelity inherent in all VR interactions. The IntFi Model considers not only the system's fidelity but also the fidelity of interactions between the user and the system because of their reciprocal relationship. Beyond physical and functional simulation of a virtual environment in the form of bits and bytes, VR technology requires accounting for how the user's body affects the virtual world and how output devices can generate physical stimuli. This makes a holistic integration of all elements of embodied 3D interactions imperative. Therefore, the IntFi Model is based on the human-computer interaction (HCI) loop [71], a well-established design principle that breaks down how a user and a system perceive and influence each other. For this, one aspect of fidelity is assigned to each of the eight stages of the HCI loop, as illustrated in Figure 1. As a result, the IntFi Model consists of eight distinct fidelity components: (i) action, (ii) detection, (iii) transfer, (iv) simulation, (v) rendering, (vi) display, (vii) perceptual, and (viii) experiential fidelity.

Building on prior work, the proposed model establishes a clear structure of the fidelity components with consistent terminology, precise definitions, detailed explanations, and illustrative examples in Section 3. This paper serves as a signpost by referring to more specialized frameworks and models detailing single components beyond the scope of this work. The IntFi Model can also help set a rigorous research agenda to advance purposeful measurement methods, determine factors contributing to fidelity, and understand the interdependence of the individual components. The IntFi Model can inform the VR community on how to focus its efforts to achieve a broad comprehension of realistic interactions. Beyond demonstrating how realistic a simulation is or how its fidelity differs from another, the model's theoretical foundation allows us to understand why [113]. Hence, theory-driven study designs facilitate more generalizable evaluation results and make linking them to other research easier.

The next section summarizes previous approaches, terms, and conceptualizations of fidelity and related constructs. We then present the IntFi Model in detail in Section 3. The subsequent section demonstrates the model's application with three example use cases. In Section 5, we describe the validation process and report the findings from 14 semi-structured, one-hour interviews with VR experts from research and the industry. In Section 6.1, we share best practices for applying the model, explain how it can serve as different lenses through which VR interactions can be viewed, and caution against common application traps. Section 7 provides a discussion of typical fidelity patterns, practical and theoretical implications, and limitations of the proposed model. Lastly, we propose an abundant research agenda that the IntFi Model opens up and can inspire future work in our field in Section 8.

2 RELATED WORK

As soon as users enter a VE, they interact with the simulation. Within the scope of this paper, we consider interaction as a reciprocal exchange of a user and a computer system observing and reacting to each other through actions and states. The actions taken by users and the output generated by a computer depend on each other and together form the interaction. Therefore, interactions with a computer system cannot be attributed solely to humans or computers. The two must be considered together [37]. The two-sided behavior happens simultaneously, continuously, inseparably, and inevitably, similar to a person affecting and being affected by the world around them in reality.

Although this exchange simultaneously occurs in both directions, it can be helpful to think about the interaction as a circular sequence of steps for conceptually distinguishing them. A software design framework that describes this circular process is the model-view-controller (MVC) pattern introduced by Reenskaug [81] in 1979. This software architecture pattern became one of the most influential for describing and developing user interfaces. We illustrate the process in Figure 3. A design principle that describes a similar process with HCI-specific labels is what we call the *human-computer interaction loop*. Forming the inner circle in Figure 1, the HCI loop links the user, the input devices, the computer, and the output devices with transitions between these elements. This model can be traced back to a chapter by David Owen in "User Centered System Design" (p. 368) edited by Don Norman and Stephen Draper [71]. In the same book, Norman describes a similar process from a cognitive perspective of how users must overcome the Gulfs of Execution and Evaluation. He argues that the user continuously evaluates the current system state and plans actions to accomplish a specific goal—input and feedback. Numerous frameworks and textbooks adapted the loop for different purposes and specializations, making it an established tool in HCI research [1, 39, 56, 59, 72, 111]. Combined, the states and transitions of the HCI loop form the eight components of the proposed IntFi Model, and the two gulfs correspond to the grouping of input into and output from the system.

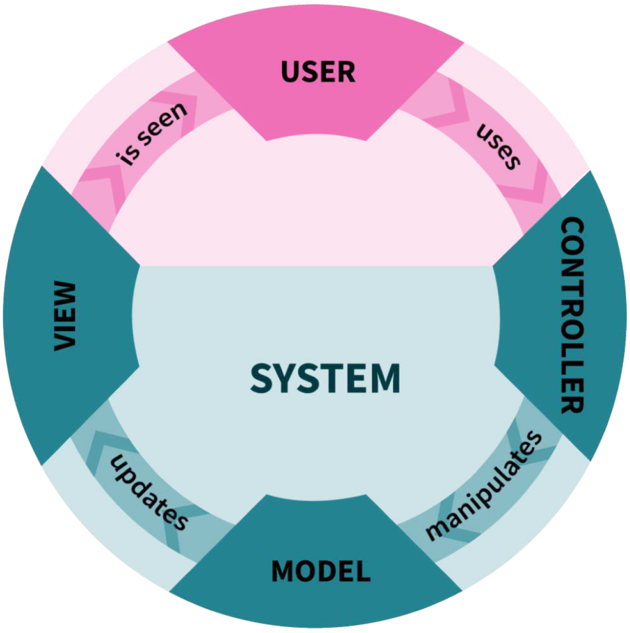


Figure 3: The Model-View-Controller pattern [81] from 1979 on which the IntFi Model is conceptually based: A user uses a controller to manipulate the model, which updates the view and is then seen by the user.

2.1 Understanding Interaction Fidelity

In the context of VR, the interaction with a system has the purpose for the user to experience and influence a simulated reality. As outlined in the introduction, the level of fidelity plays a vital role in many VR simulations. Several frameworks, models, and evaluations have investigated various aspects of interaction fidelity. Please note that we refer to the original terms from cited works instead of the IntFi Model's terminology in this subsection. As a result, they might appear confusing and contradictory.

Most prominently, McMahan [55] proposed the *Framework for Interaction Fidelity Analysis* (FIFA) that was released in an updated version in 2016 [59]. The revised FIFA considers the three categories *biomechanical symmetry* describing the reproduction of body movements from the real world, *input veracity*, which considers the exactness of input devices capturing movements, and *control symmetry*, which covers the exactness of control in the virtual world compared to the real world. Each category comprises further detailed components. The framework is designed to compare the user's motions during virtual activities that involve body movements, such as techniques for locomotion or object manipulation, to their counterpart in reality. Results from their user studies suggest an uncanny valley of VR interactions. They found good user performance with low- and high-fidelity systems but a drop in performance with medium-fidelity systems. Similar findings have been presented in further studies [12, 66]. In other investigations, Bowman et al. [16] also found that high-fidelity interactions can enhance performance and the overall user experience, but medium levels of fidelity can be unfamiliar and detrimental to performance.

The FIFA framework, however, only considers the fidelity of user actions and, therefore, only the input side of the two-way interaction. The system output is neglected in this framework even though it is an inseparable part of the interaction and can heavily influence the realism of a system. McMahan et al. [58] acknowledge the missing output component by also analyzing the display fidelity and finding a similar negative effect for medium display fidelity as for their interaction fidelity. Nilsson et al. [67] build on this finding and argue that when some fidelity components are limited, maximizing the fidelity of other components may be detrimental to the perceived realism. Therefore, decreased fidelity might positively influence perceived realism in some instances. In addition, Abtahi et al. [2] show that even for interactions that go beyond reality, some aspects of the interaction should still be grounded in the real world to avoid sensory conflicts in the user.

For evaluating the fidelity of a VR simulation, Stoffregen et al. [104] consider the action fidelity as the relationship between performance in the simulator and performance in the simulated system, the system's output in the form of optic, acoustic, mechanical, and inertial arrays, as well as the experiential fidelity in the form of perceived presence [94]. With a focus on more practical aspects of current VR systems, Al-Jundi and Tanbour [5] present a framework for evaluating fidelity concerning four interrelated elements: digital sensory system fidelity, interaction system fidelity, simulation system fidelity, and integration among these aspects to produce high-fidelity virtual experiences. They identify various factors for evaluating VR hardware regarding visual, auditory, and haptic feedback, the tracking system, and graphic quality. Conversely to the FIFA framework, this framework focuses on the output side of interactions and neglects the user actions as part of the interaction with the VR system. The fidelity of haptic feedback can also be assessed in more detail with the Haptic Fidelity Framework by Muender et al. [64], providing detailed factors to analyze and quantify aspects of sensing, hardware, and software.

In the context of gaming, Rogers et al. [83] evaluated interaction fidelity for object manipulation and whole-body movements and found that high fidelity is preferred for object manipulation. Still, moderate fidelity can suffice for whole-body movements as there is a trade-off between fidelity, usability, and social factors. Further, Rogers et al. [84] provide an in-depth analysis of realism in digital games, including a focus on VR. The authors present a two-part framework of realism dimensions consisting of a hierarchical taxonomy of realism dimensions and the mapping of realism dimensions within Adams' game model [3]. Alexander et al. [7] investigated the effect of fidelity on the transfer of knowledge from games and simulations to the real world. They argue that the fidelity of a simulation is a significant factor in enabling skill transfer and define three categories of fidelity: Physical fidelity is the degree to which the simulation looks, sounds, and feels like the real world; functional fidelity is the degree to which the simulation acts like the real world; and psychological fidelity is the degree to which the simulation replicates the psychological factors (e.g., stress, fear) experienced in the real world.

The *Reality-Based Interaction* framework by Jacob et al. [38] provides four themes to enable high-fidelity interactions on a more general level with interfaces such as touchscreens, tangibles, and VR. Interaction designers should consider naïve physics, body awareness and skills, environment awareness and skills, and social awareness and skills. The work outlines trade-offs between realism and expressiveness, efficiency, versatility, ergonomics, accessibility, and practicality. In contrast to most other frameworks that cover input and output components of the interaction, Lindeman and Beckhaus [48] focus on experiential fidelity, enhancing the realism of the user experience by guiding the user's frame of mind in a way that their expectations, attitude, and attention are aligned with the VR experience.

2.2 Inconsistent Fidelity Terminology

To this point, we have adhered to the terminology originally used in the mentioned works. The literature established a patchwork of different but similar terms based on different interpretations and assumptions. This is why the wording of the above explanations might sound inconsistent and contradictory. It demonstrates how critical uniform designations are for research communication.

The terms fidelity, realism, and naturalness were often used synonymously in previous literature. Researchers often investigated only a specific part of interaction fidelity but referred to it universally as (interaction) fidelity. For example, the term *interaction fidelity* has been used to refer to visual render quality [53], camera views and gravity [12], or dialogue capabilities [19]. Some publications refer only to the user's system input with it [16, 59]. This neglects half of the two-way interaction between the user and the system, which can only be considered in its reciprocal dependence, as outlined at the beginning of this section. Also, the literature generally refers to other individual aspects of the interaction as *fidelity*. For example, some fidelity conceptions focus on the simulated virtual environment, such as in game research [7, 50], or are reduced to the simulation's physical and functional dimensions [33, 34], while it is crucial for VR and 3D interfaces also to consider the means of input and output as well as the user's role. A recent framework classified the fidelity of mixed-reality prototyping [22]. Furthermore, outside computer science, fidelity has been narrowly defined within the fields' contexts, such as in health and psychology regarding realistic psycho-behavioural and affective responses [8, 32]. These examples illustrate how divided the VR community has been about the term's understanding and usage.

In a systematic review of the concepts of realism and fidelity for digital games, Rogers et al. [84] found a "substantial potential for confusion given the overlapping and contradictory use of realism types." The authors report that the type of realism is often not even further defined but remains vague in the literature. The rigorous analysis covers VR research as part of gaming but excludes the realism of other VR interactions and the fidelity compared to other reference frames. Nevertheless, the survey outlines the vast range of terms used to describe aspects of realism and fidelity. This emphasizes the urgent need for a theoretical basis of consistent terminology. Plenty of research contributes to the understanding of the multidimensional concept of fidelity. Still, it lacks an umbrella model into which the individual elements can be integrated to understand the bigger picture. We will consequently use the IntFi Model's terminology for the remainder of this paper.

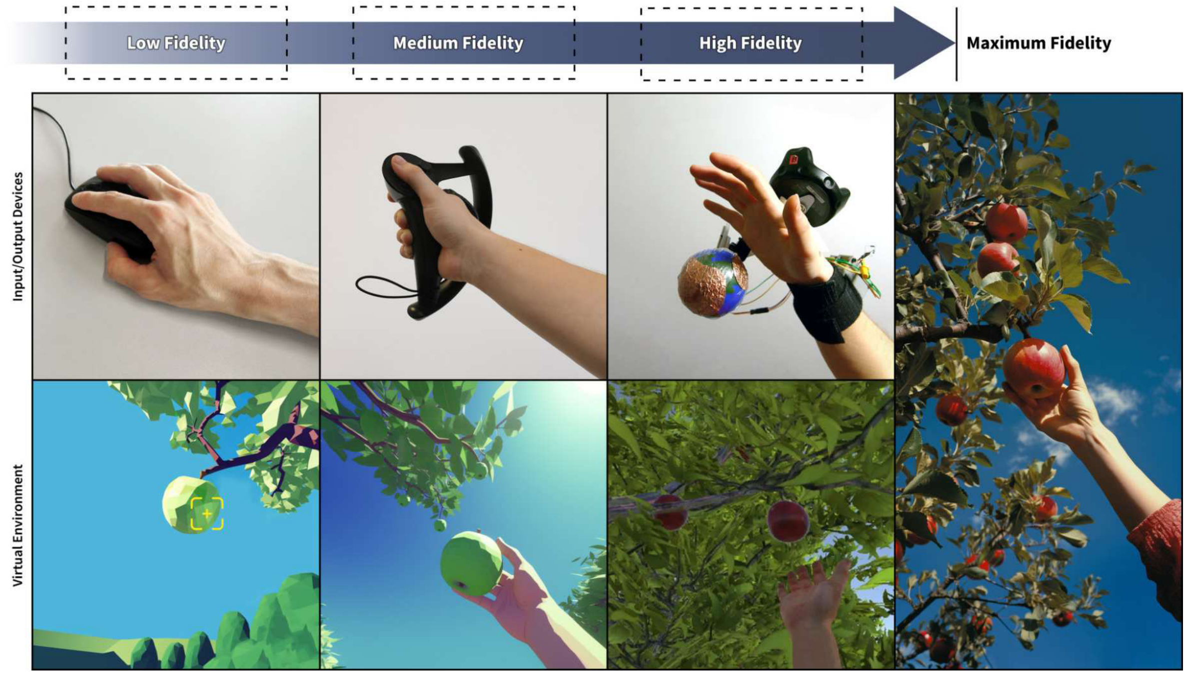


Figure 4: The fidelity spectrum with approximate ranges from low, medium, and high to maximum fidelity with an example use case: implementations with different fidelity levels of somebody picking an apple from a tree. The reference interaction from the real world on the right side is defined as maximum fidelity. ***Copyright of the "High Fidelity" images by de Tinguy et al. [23]. The other images have been photographed or generated with Midjourney by the authors.***

3 MODEL OF INTERACTION FIDELITY

The conceptual model presented here distinguishes various aspects of the fidelity of interactions in VR. It covers the entire process of a user interacting with a VR system, from user input over system processing to output from the system experienced by the user. Instead of assessing the contribution of each device or system component to the fidelity of the interaction, we propose distinguishing between the stages of the interaction to systematically evaluate how true it is to the original. The IntFi Model is based on the HCI loop [71], which originates from the model-view-controller pattern described in Section 2. Following the structure of the loop, the model assigns one aspect of fidelity (for example, display fidelity) to one stage of the loop (in this example, output devices), as illustrated in Figure 1. The loop offers simplicity, yet all fidelity aspects of any conceivable interaction are integrated. Therefore, it is a sound foundation for the intuitive differentiation of factors that define the fidelity of any VR interaction with the user in mind.

Based on the Merriam-Webster Dictionary [63], McMahan [55], Alexander et al. [7], and Raser [79], we consider *interaction fidelity* as the degree of exactness with which reference interactions are reproduced. Thus, it describes how closely a user's interactions with a VR system resemble the interactions from a reference system. This reference system can be the real world, in which case we refer to realism, but we can also choose any other reference interaction, such as fictional worlds (e.g., Star Wars), hyper-realistic interaction techniques (e.g., the Go-Go technique [76]), a previous VR system, a planned system iteration, or a replicated study. It is important to clearly define the chosen reference interaction for a meaningful and unambiguous fidelity assessment. If the reference is changed to make another comparison, the assessed fidelity will also change.

We can describe the level of fidelity on a spectrum covering low, medium, and high to maximum fidelity as illustrated in Figure 4. With maximum fidelity, there is theoretically a perfect correspondence to the original, even if it might be technologically impossible to achieve. Between perfect and no correspondence, there is a continuum [12, 15, 22, 50] without clear-cut "low", "medium", or "high" states. This wording demonstrates a relative difference or approximate range on the continuum.

It is crucial to keep in mind that fidelity is an objective concept simply describing the degree of correspondence without judgment. Higher interaction fidelity is not necessarily better, more desirable, more effective, or more immersive but merely implies a closer match to the reference. Although higher fidelity can have benefits for other metrics or goals, it has also been shown how lower-fidelity and hyper-natural interactions can be beneficial [25, 26, 35, 57, 60, 66]. On the other hand, aspects of fidelity often determine the success of a simulation. For instance, in motor skill learning, the faithfulness of the user's movements is crucial. Likewise, authentic scenic details are the key aspect of a travel simulation, accurate haptic feedback during surgical training, and plausible situations for phobia therapy. As Section 7 outlines, objective system fidelity does not necessarily correlate with perceived realism. For effective and economical planning, interaction designers and VR developers must reflect on what kind of fidelity is important for the use case.

All aspects of fidelity can be assessed objectively and subjectively depending on the point of view. When applying standardized metrics for reproducible, indisputable measures to describe fidelity, we can objectively determine and verify the exactness of the interaction's match with the reference interaction. For example, we can impartially compare two screens regarding their technical specifications, such as pixel density. When relying on personal impressions from interviews or questionnaires, we can subjectively assess fidelity. For example, we can ask a user in a questionnaire how closely the hand movements in a VR juggling training experience match juggling in reality. Some fidelity aspects can be assessed objectively in a more meaningful way, such as system specifications (e.g., screen resolution) or historical facts (e.g., the 1988 ACM Turing Award recipient) with verifiable ground truth. For other aspects, on the other hand, it might make sense to assess them subjectively, such as perception or experience. Currently, we do not have the means to determine every aspect objectively. This might be technically possible in the distant future, even for experiential fidelity, with sufficiently sophisticated brain–computer interfaces.

One could argue that users' perceptions and experiences are inherently subjective because they vary between individuals. However, while they are different between users, we can assess perceptual and experiential fidelity individually: How would the same person probably perceive and experience the reference interaction? Because systems are usually not tailored to individuals, typical or average users from a target group can be considered for better generalizability. For this pragmatic reason, we advocate for a population-centric assessment through user-centric research. Some fidelity aspects can be assessed independently (e.g., rendering fidelity), but especially for the user-related aspects (i.e., perceptual, experiential, and action fidelity), the target users' abilities and characteristics must be considered to provide accessible systems acknowledging the diversity of users. For example, people with color vision deficiency perceive the same visual output differently, which might affect how closely it resembles their real-life perception. Similarly, target populations can experience a system's fidelity differently, for example, depending on their expertise and how competent they feel in a virtual experience compared to their real-world competence. For example, experienced soccer players feel more restricted than novices in virtual kicking with medium simulation fidelity [13]. In discourse, we must keep in mind that we all perceive the world subjectively and create a mental model of how it works [70]. While people can have different perspectives on theoretical ground truth and be challenged in their view, we need to agree in discussions on an explicit reference interaction that is supposed to be simulated and target user groups for meaningfully applying the term fidelity.

3.1 Development of the Model

We have devised the idea for the IntFi Model when gathering different aspects of fidelity from the literature and testing possible classifications to bring structure to the concept. When cross-referencing the first approaches with related work, we realized that any dimension fits neatly into the HCI loop. In an iterative process, we refined the labels and definitions of the components from discussions among the authors, with research peers, in teaching practice, and at conferences. We conducted semi-structured expert interviews with 14 VR researchers and practitioners to improve and validate the model. We present the method and results in detail in Section 5.

3.2 Structure of the Model

The model consists of the user and the VR system, which includes input devices, the computer as the processing unit with data and models, and output devices. Between these components, Figure 1 uses arrows indicating a translation from software to hardware (such as the rendering from the simulation to the output devices) or, vice versa, from physical to intangible information (such as from the system output to the user's mind through perception). Every aspect of interaction fidelity corresponds to one stage of the HCI loop and is, thus, represented by one component in the model, as visualized in Figure 1. For example, the fidelity with which the system detects the user actions corresponds to the *input devices* in the HCI loop, which is the *controller* in the MVC paradigm and is linked to the *detection fidelity* in this model. The single components of the model are further detailed in this section.

We can consider the model vertically and distinguish aspects of *input fidelity* (right side) and *output fidelity* (left side). We can also consider the model horizontally and distinguish aspects of fidelity that concern the *user* with their perceptions, experiences, and actions (upper part) and those concerning *system fidelity* with the detection, transfer, simulation, rendering, and displays (lower part). Input fidelity in this model is close to what McMahan [55] described as interaction fidelity in the Framework for Interaction Fidelity Analysis (FIFA) and comprises similar components: *action*, *detection*, and *transfer fidelity*. All further components were not considered in FIFA.

Aspects of fidelity that determine the characteristics of the VE and react to the user input are included in the component *simulation fidelity*. Proceeding in the loop, the group of *output fidelity* components comprises *rendering fidelity*, *display fidelity*, and *sensory fidelity*. Finally, all these aspects combined determine *experiential fidelity*, the impression created in the user's mind. In the IntFi Model, we focus on the endpoints of the single components instead of elaborating on the technical processes behind each component. For example, detection fidelity is determined by the final output of the input device's API, not by the sensor's firmware or signal processing. The single components can be broken down further as needed (e.g., interaction fidelity → simulation fidelity → presentational fidelity → 3D model → skin texture → height map → resolution). In the scope of this paper, we will only detail conceivable subcomponents through examples, not comprehensively, except for *simulation fidelity*. As a whole, all aspects of the model define overall interaction fidelity. In the following, the components are explained in detail following this structure: We define the fidelity aspect, distinguish it from the subsequent aspect, detail its characteristics, state its requirements for maximum fidelity, illustrate how different levels of fidelity could be designed, and refer to specialized frameworks or similar definitions.

Table 1: The definitions of the fidelity components and the categorizing sets of components.

Aspect	is defined as the degree of exactness with which...	Depends on
Action Fidelity	user actions resemble those of the reference interaction.	User
Detection Fidelity	input devices detect the user's actions.	System
Transfer Fidelity	virtual actions, derived from the input measurements, resemble the user's actions of the reference interaction.	System
Simulation Fidelity	a virtual environment resembles the characteristics of the reference interaction's world and adequately reacts to the user's actions.	System
Rendering Fidelity	the output content generated by the computer resembles what would be presented to the user in the reference interaction.	System
Display Fidelity	the output devices reproduce the physical stimuli presented to the user in the reference interaction.	System
Perceptual Fidelity	the user's perception of the physical stimuli created by the system resembles how the user would perceive the reference interaction.	User
Experiential Fidelity	the user's experience of the simulated interaction resembles how the user would experience the reference interaction.	User
Set of aspects	**is defined as the degree of exactness with which...**	**Includes**
Input Fidelity	the virtual actions generated from the user's input resemble the user actions of the reference interaction.	Action, Detection, Transfer
Output Fidelity	the system output is generated and perceived as the user would perceive it in the reference interaction.	Rendering, Display, Perceptual
System Fidelity	the system reproduces the world of the reference interaction reacting to the user.	Detection, Transfer, Simulation, Rendering, Display
Interaction Fidelity	**reference interactions are reproduced.**	**All aspects**

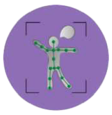
3.2.1 *Action Fidelity*.

Action fidelity is the degree of exactness with which user actions resemble those of the reference interaction. In the context of this model, we consider actions as any active behavior or passive state of the user, including sheer existence. In contrast to the next component, *detection fidelity*, it is irrelevant for action fidelity whether the system captures the user actions.

This component comprises all behavior and states of the user, including body movements, such as grasping or walking, and any other modality and activity, such as speaking, eye gaze, facial expressions, or brain activities. It is crucial to consider the full range of user actions. Not only intentional actions (e.g., gestures) are relevant for action fidelity but also subconscious (e.g., blinking), uncontrolled (e.g., blushing, swelling), involuntary (e.g., tremor), passive (e.g., static poses), and unaware actions (e.g., body temperature).

In the example of grasping an object, the user of a low-fidelity solution would point at the object with a 3-degrees-of-freedom (DoF) controller and hold it by pressing a button. With high action fidelity, the user would reach out the hand to the object's position, enclose it with the fingers according to its shape and size, and exert force with the arm in proportion to its weight. To virtually reproduce interactions with maximum fidelity, the user must behave exactly as in the reference interaction.

This component is close to the *biomechanical symmetry* dimension of the revised FIFA by McMahan et al. [59]. In their definition, only the movements of body parts are considered user actions. They distinguish between anthropomorphic, kinematic, and kinetic symmetry of body movements. We include additional forms of user actions in this model beyond active bodily motion.



3.2.2 ***Detection Fidelity***. Detection fidelity is the degree of exactness with which input devices detect the user's actions. In contrast to the next component, *transfer fidelity*, it is irrelevant for detection fidelity how the system interprets or classifies the captured sensor data. While detection fidelity is about converting physical actions into measurements of those actions, transfer fidelity is about converting those measurements into meaningful virtual actions.

This component involves the capability of sensors to capture relevant signals. For high detection fidelity, the measurements are accurate, noiseless, reliable, and immediate. The kind of signal and the appropriate input device depend on the respective user actions and concerns the same range outlined under action fidelity, from optical tracking of body movements to microphones for speech detection to electromyography (EMG) for facial expressions. The sensor range can cover all possible parameters, such as thermosensors for measuring body temperature. The system does not need to detect actions or body signals that would not affect the reference interaction, such as the temperature of the feet when knitting with one's hands.

For example, a low-fidelity solution in a horror game would detect the user jumping in terror and gasping. In contrast, in a high-fidelity solution, the system would also detect the user's eyes snapping open, their muscles contracting, and the intensity of their sweating. To virtually reproduce interactions with maximum fidelity, the input devices must detect every relevant user action and state precisely.

This component is close to the *input veracity* dimension of the revised FIFA [59]. The authors distinguish between the measurement's accuracy, precision, and latency as the delay before sensory feedback. We consider the immediacy of the detection as a decisive factor but independent of system feedback.

3.2.3 ***Transfer Fidelity***. Transfer fidelity is the degree of exactness with which virtual actions, derived from the input measurements, resemble the user's actions of the reference interaction. In contrast to the next component, *simulation fidelity*, it is irrelevant for transfer fidelity how the virtual actions affect the simulation.

This component refers to how the system considers the sensor readings, processes and transforms them, and interprets them in the context of the simulation to generate virtual actions mapped from the user's real actions in a meaningful way. The virtual actions can deliberately differ from the real actions, e.g., by modifying the control/display ratio. Therefore, this component is determined by the correspondence of the virtual actions with the reference interaction to be simulated, not with the actual user actions, if these differ. For example, in redirected walking [80], we compare the virtual path to the intended straight path, not the circular path that the user physically takes. In case of incomplete, distorted, or simplified input data, processing can make up for input deficiencies to allow finding a probable interpretation.

As an example of low transfer fidelity, when processing noisy capacitive sensor data of a hand-held controller for approximating the hand pose, a low-fidelity solution would show jittery and anatomically absurd finger movements. In contrast, a high-fidelity solution would result in smooth and plausible finger movements that match the real hand pose of the user, e.g., using inverse kinematics. To virtually reproduce interactions with maximum fidelity, the transfer function must allow the system to correctly interpret the measurements, infer the correct meaning, and enable the virtual actions to affect the simulation appropriately.

This component is close to the *control symmetry* dimension of the revised FIFA [59]. It is described to only depend on the transfer function symmetry where the system's transfer function is contrasted with a theoretical transfer function from reality. While we agree that no transfer function that would correspond to this system component exists in reality, we consider matching the virtual replication of the user actions with its counterpart of the reference interaction more practical than constructing a theoretical dummy.

3.2.4 ***Simulation Fidelity***. Simulation fidelity is the degree of exactness with which a virtual environment resembles the characteristics of the reference interaction's world and adequately reacts to the user's actions. In contrast to the next component, *rendering fidelity*, it is irrelevant for simulation fidelity how the VE is rendered as output. While simulation fidelity considers the environment with its characteristics, rendering fidelity converts those characteristics through a dynamic process to the output devices (i.e., rendering pipeline).

This component concerns all elements of the simulated environment, including objects, agents, physics, and any other entity or logic from the reference interaction's world. To help categorize these aspects and to avoid simulation fidelity being a "black box", we define four subcomponents: presentational fidelity, behavioral fidelity, physical fidelity, and scenario fidelity.

Presentational fidelity is the degree of exactness with which the presentation of the simulated world resembles the reference world. Properties that can affect presentational fidelity include an object's mesh, colors, materials, scent, surface, reflectivity, and level of detail. For example, a low-poly presentation of an apple with a simple color-based material would have lower presentational fidelity to a real-world apple than a laser-scanned presentation with a photorealistic texture-based material.

Behavioral fidelity is the degree of exactness with which the behaviors of agents (e.g., virtual humans) within the simulation resemble the behaviors of their counterparts in the reference interaction's world. The fidelity of these behaviors depends on three aspects of the agent: its perceptual model, cognitive model, and motor model [56]. The perceptual model defines what information is made available to the agent, which can vary from all information about the state of the world (i.e., fully observable) to a subset of information (i.e., partially observable) [86]. The cognitive model defines how the agent will generate actions based on the information provided by the perceptual model, such as simple reflexive responses or more complex decisions based on prior events and the agent's goals. The motor model defines what observable actions the agent can control, such as body movements, facial expressions, speech, or manipulating objects within the environment. Hence, a simple reflex agent that chooses actions from a small set based on current, partially observable information would have less behavioral fidelity than a goal-based agent that chooses actions from a large set of possibilities given fully observable information about the simulation.

Physical fidelity is the degree of exactness with which the physics of the simulation resembles the physics of the reference interaction's world. Physical fidelity can be affected by shapes, mass, distribution of mass, drag, gravity, or whether collisions are calculated discretely or continuously. For example, when skipping a stone over water, a high-fidelity system would simulate how the stone bounces due to the surface tension, travels a realistic trajectory, and creates ripples on the water's surface. In contrast, a stone with low physical fidelity (such as a box collider, a default mass of 1kg, and simple force estimations) would simply sink to the ground.

Scenario fidelity is the degree of exactness with which the simulated situation resembles the situation of the reference interaction. While presentational and behavioral fidelity focus on individual objects or agents, scenario fidelity focuses on the holistic aspects of the simulation. This involves the spatial relationships among objects (e.g., a chair is usually placed under a table and not on top of it), their semantic relationships (e.g., an indoor room usually has a ceiling), and their logical relationships (e.g., flipping a light switch turns the light on or off). For example, a VE with a flat ground plane and no vegetation would provide less scenario fidelity to hiking through a mountain forest than an environment with uneven terrain and numerous trees.

As an example of different simulation fidelity, consider the reference interaction of playing tennis against a friend at your local park. A low-fidelity implementation may use low-poly representations of the tennis ball, rackets, and net with a basic virtual agent as the opponent that only reacts to the user hitting the ball, which follows a simple trajectory. On the other hand, a high-fidelity implementation would use high-poly representations in an outdoor park environment surrounding the tennis court, a virtual human driven by its own goals and capable of social interactions (e.g., congratulatory comments), and advanced physics that allows the user to apply topspin when hitting the ball. To virtually reproduce interactions with maximum fidelity, the simulation must replicate the reference world identically, including its objects, agents, physics, and situation.

The first part of the definition is based on McMahan [55]. This component closely links to the concept of simulation fidelity by Nilsson et al. [67]. They accept the definition by McMahan [55] and further attribute simulation fidelity to "the realism of the models forming the basis for the generation of the VE (e.g., geometric, lighting, or physical models)." While we largely agree, we extend the scope of simulation fidelity by the system's response to the user's input. This is also reflected in the concept of functional fidelity by Alexander et al. [7], which requires in a training context that "the simulation acts like the operational equipment in reacting to the tasks executed by the trainee." However, in this definition, the characteristics of the VE independent of the task execution are disregarded. Further, we adopt and expand the subcategories *attribute, behavioral, and physical coherence* by McMahan [56] as well as *physical and functional characteristics* by Hays and Singer [34].

3.2.5 ***Rendering Fidelity.*** Rendering fidelity is the degree of exactness with which the output content generated by the computer resembles what would be presented to the user in the reference interaction. In contrast to the next component, *display fidelity*, it is irrelevant for rendering fidelity whether the output devices can display the rendered output (accurately) and make it perceptible to the user.

This component involves any modality, including visual, auditory, haptic, olfactory, gustatory, and vestibular stimuli. The sensory stimuli must be rendered for any output device of the system. Systems with a visual display must render graphical images that can differ in their resolution, aspect ratio, framerate, visual style, antialiasing, texture resolution, detail of height maps, shadow, specular effects, and much more. Similarly, the sensory stimuli for any other kind of display must be rendered considering the respective parameters affecting the level of fidelity, such as audio, haptics, etc. (see *display fidelity*). The rendering parameters do not necessarily correspond to the parameters of the simulation or the display; e.g., the rendered audio might compress the audio sources from the simulation and still have a higher resolution than what the earphones can display.

As an example, for the sound of a virtual human walking, a low-fidelity solution would play a loop of the same footstep recording with strong compression. In contrast, a high-fidelity solution would synchronize the timing with the foot movement, adjust the volume, pitch, reverberation, and direction to the user's position, and dynamically blend high-resolution sound samples matching the floor material and the impact of the foot. To virtually reproduce interactions with maximum fidelity, the rendered output for all modalities must be indistinguishable from what the user would perceive in the reference interaction.

This component is related to the definition of display fidelity by McMahan et al. [58]: "the objective degree of exactness with which real-world sensory stimuli are reproduced." In our model, we divide this aspect into rendering and display fidelity to reflect the independence of calculating output from displaying it.

3.2.6 ***Display Fidelity.*** Display fidelity is the degree of exactness with which the output devices reproduce the physical stimuli presented to the user in the reference interaction. In contrast to the next component, *perceptual fidelity*, it is irrelevant for display fidelity whether and how the user perceives the stimuli.

This component covers displays that concern any modality, including visual, auditory, haptic, olfactory, gustatory, and vestibular displays. Since the alignment of output devices and human perception is highly complex, countless factors must be considered for display fidelity. For instance, concerning the graphical output of a head-mounted display, visual fidelity can be affected by the screen resolution, pixel density, field of view, refresh rate, contrast, and color depth, but also optical properties of the lenses and optical interferences such as god rays. The situation is again entirely different in a CAVE system with wall projections. With any modality, the properties of the displays might not match those of the renderings or what the user would be capable of perceiving, e.g., a screen might have a different resolution than the rendered image and the user's retina.

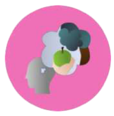
As an example of haptic fidelity, considering the task of sawing through a plank, a low-fidelity solution would create static vibration feedback in a hand-held controller when moving the saw. In contrast, a high-fidelity solution would display force feedback that resembles the resistance of the wood and restricts the hand's lateral movement, as well as generate contact forces from the saw handle on the hand, render the gravitational pull of the heavy saw, and produce dynamic vibrations matching the jerky movement through the wood. To virtually reproduce interactions with maximum fidelity, the sensory stimuli displayed to the user address all senses and perfectly resemble the stimuli in the reference interaction.

This component is related to various frameworks that detail display fidelity addressing a single sense, such as the Haptic Fidelity Framework by Muender et al. [64], components of visual display fidelity according to Bowman and McMahan [15], or dimensions of the Spatial Audio Questionnaire by Lindau et al. [47].

3.2.7 ***Perceptual Fidelity***. Perceptual fidelity is the degree of exactness with which the user's perception of the physical stimuli created by the system resembles how the user would perceive the reference interaction. In contrast to the next component, *experiential fidelity*, it is irrelevant for perceptual fidelity how users interpret the perceived stimuli, what meaning they assign to them, and what consequences they draw.

This component concerns all sensory cues that the user registers with any sense: vision, audition, touch, smell, taste, proprioception, equilibrioception, nociception, etc. The physical stimuli produced by the output devices are registered through sensory receptors. The information is transduced to and processed in the user's brain. The impressions from the different senses are integrated into one unified perception, called multimodal integration. We consider the interpretation of the stimuli already part of the user's experience and, therefore, included in experiential fidelity. We assume that maximum display fidelity inevitably leads to maximum perceptual fidelity because the human sensory system cannot identify the origin of a physical stimulus, whether the real world or a simulation generates it. With an imperfect display, however, perceptual fidelity may deviate. Mechanisms in human perception can compensate for display deficiencies, enabling phenomena such as illusions, biases, or sensory substitution. Thus, high perceptual fidelity may be achieved even without high display fidelity in some cases. Further, the qualia of perception can vary immensely between individuals, which is why perceptual fidelity must be assessed on a user-by-user basis. Consider a color-blind user who experiences the real world in shades of grey. A black-and-white scene would look more realistic to that user than to users with full-color vision, thus yielding higher perceptual fidelity. Considering future possibilities of perception through direct neural manipulation without the respective receptors being stimulated, our understanding of this component remains the same: How closely does the user's perception correspond to the reference? However, the restriction to simulation through physical cues must then be omitted.

In the example of someone touching different locations on the user's back, a solution with low display fidelity, such as a haptic vest with low-resolution actuators, would still result in high perceptual fidelity, as the two-point discrimination of skin receptors on the back is relatively poor [46]. A low-fidelity solution would have the same display resolution on the user's hand as the receptors are more sensitive to local variations here. Perceptual fidelity is also low in this example if the user only perceives visual cues and feels no tactile sensation. To virtually reproduce interactions with maximum fidelity, the stimuli by the output devices must evoke the same sensation for all senses and create the same perception in the user's brain as in the reference interaction.



3.2.8 ***Experiential Fidelity***. Experiential fidelity is the degree of exactness with which the user's experience of the simulated interaction resembles how the user would experience the reference interaction. When comparing to the real world, this is often referred to as *perceived realism*. In contrast to the next component, *action fidelity*, it is irrelevant for experiential fidelity how the user reacts to the simulation.

This component is based on how faithful to the reference interaction the user considers their own actions, how the observed world behaves, and how it can be perceived. Experiential fidelity is often the ultimate objective when optimizing any other component, as the perceived authenticity can determine the subjective quality of a simulation. In other use cases, however, it can be subordinate, as in a technical proof of concept or a training situation that must prioritize action or simulation fidelity, whether experienced as faithful or not. Curiously, an objectively highly realistic system is not necessarily experienced as such, as discussed in Section 7.6. Various factors affect experiential fidelity, such as individual differences in perception and judgment, since systems are often not optimized for one specific user and use case but are designed to be versatile and adaptable. Also, suspension of disbelief—or lack thereof—can substantially impact perceived fidelity, e.g., due to different assumptions and expectations, distractions from other realities, and the credibility or plausibility of the simulation. Also, unconscious effects need to be considered. For example, users can experience a higher cognitive load even for unnoticeable manipulations with redirected walking techniques [17]. Furthermore, multisensory integration can influence how users interpret conflicting stimuli, which can be demonstrated with phenomena such as the McGurk effect [107]. Another critical factor is the awareness of the experience being simulated or the memory of having entered a simulation.

For example, when interacting with highly but not perfectly realistic virtual humans, the experiential fidelity has often been found to be low due to the uncanny valley [54], despite high rendering and display fidelity. On the other hand, even with low display fidelity, it is possible to achieve high experiential fidelity, e.g., Valve's experience of drawing a longbow with only vibration and sound

Table 2: Selected frameworks, models, and instruments that include details specialized on a certain aspect of fidelity, listed with its original designation. The references are grouped by their correspondence to the IntFi ModelṪhis does not represent an exhaustive list of all relevant, prior literature, nor does it comprehensively address all subcomponents.

IntFi Model components	Literature covering this component
Action Fidelity	*Biomechanical Symmetry* [55] (only active motions), *Motion Realism* [84] (only in games)
Detection Fidelity	*Input Veracity / System Appropriateness* [55], *Tracking System Fidelity* [5]
Transfer Fidelity	*Control Symmetry* [55]
Simulation Fidelity	*Simulation System Fidelity* [5], *Physics Realism / Avatar Realism* [84] (only in games), *Functional Fidelity* [7], Naive Physics [38]
Rendering Fidelity	*Software* [64] (only haptics), *Visual/Graphic & Auditory Realism* [84] (only in games)
Display Fidelity	*Hardware* [64] (only haptics), *Spatial Audio Quality* [47] (only audio), *Visual & Auditory & Haptic System Fidelity* [5], *Device Realism* [84] (only in games)
Perceptual Fidelity	*Sensing* [64] (only haptics), *Spatial Audio Quality* [47] (only audio), *Sensory Realism* [84] (only in games), Body Awareness and Skills [38]
Experiential Fidelity	*Experiential Fidelity* [48], *Player Response Realism* [84] (only in games), *Presence* [94, 100], *Psychological Fidelity* [7], *Haptic Experience (HX)* [42], Environment & Social Awareness and Skills [38]
Input Fidelity	*Action Fidelity* [104]
Output Fidelity	*Physical Fidelity* [7], *Digital Sensory System Fidelity* [5]
System Fidelity	*System Fidelity* [56]

feedback using a hand-held controller.[1] Another example is the rubber hand illusion and its virtual replication [51], which demonstrates how users can have strong body ownership and perceive haptic sensations with high perceived realism despite only sensing visual cues. To virtually reproduce interactions with maximum fidelity, the interaction with the system must convince the user to experience the reference interaction, not a simulation. This would be the equivalent of a successful Turing Test for VR interactions, as proposed by Stoffregen et al. [104].

This component is related to concepts such as presence, immersion, coherence, or body ownership and is assessed in several corresponding questionnaires [94]. We discuss this in more detail in Section 7.3.

3.3 Dedicated Literature on Subcomponents

With these eight distinct components, the IntFi Model illustrates what design decisions influence the fidelity and characteristics of VR systems enabling immersive interactions. When zooming out, the model provides an umbrella framework. When zooming in, further specialized models are needed to investigate the underlying complexity of the components. In multi-modal simulations, fidelity has countless detailed determinants that can be finely dissected as required. For example, when looking at the wrinkles of a virtual human, we can go further down in the component hierarchy: interaction fidelity → simulation fidelity → presentational fidelity → 3D model → skin texture → height map → resolution. But while skin characteristics can be rendered visually, they can also be rendered haptically: interaction fidelity → display fidelity → haptics. The Haptic Fidelity Framework [64] is a good example that illustrates the complexity of one of the modalities of display fidelity. The framework comprises 14 distinct criteria defining just this one output modality.

In the interest of this model's simplicity, we refrained from further detailing the included components. Instead, the work builds on various rich and informative works we refer to in Table 2 as a signpost. It lists related frameworks and models from the literature that tie into the components of the IntFi Model. They specialize in one or a few aspects and provide in-depth information as needed. While the referred works provide further details concerning a component, they are not necessarily in exact correspondence with the component. Beyond the works listed, we encourage the HCI and VR community to devise further dedicated frameworks and measurement instruments to fill the current gaps. For instance, regarding simulation fidelity, the broad range of influences is not yet covered adequately by any specialized framework. Further, while the FIFA framework [59] allows a detailed analysis of body movements as a part of input fidelity, other elements of user actions and states are disregarded, such as speech, gaze, or body temperature.

4 EXAMPLES OF APPLYING THE MODEL

Let us look at three diverse examples to bring the theory to life. In this section, we walk you through the analysis process of three use cases with different types of reference interactions. Example 1 demonstrates how we can use the IntFi Model to evaluate how realistic a training system for surgeons is. The goal is to come as close as possible to real surgery to practice under safe conditions.

[1] The Lab (https://steamcommunity.com/app/450390?, last access: 2023-08-15). Valve Corporation, 2016.

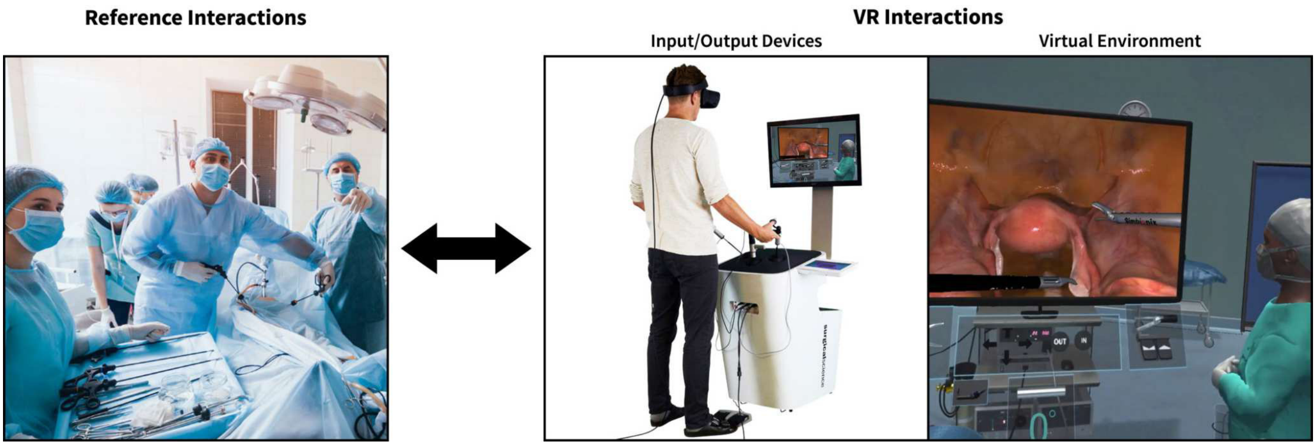


Figure 5: (left) The reference interactions in the real world to be simulated: A surgeon and his team performing a laparoscopy. (middle) The user operating the VR training system. (right) The user's view inside the VE: the operating room and a virtual screen with the laparoscopic camera view. ***Images by Surgical Science, modified by the authors.***

Therefore, high interaction fidelity is essential for acquiring skills in VR so they can be transferred to real-life surgery. In contrast, Example 2 outlines skill training in which low-fidelity elements are helpful. Here, we assess the realism of a game for learning to juggle. The goal is to provide deliberately low interaction fidelity for effective training, accompanied by empirical research [4]. Example 3 shows how the IntFi Model can be applied to fictional reference interactions. We discuss the fidelity of a VR game from the *Star Wars* universe.

The authors conducted these exemplary assessments. They are subjective in nature, thus contestable. The examples illustrate how a complete evaluation based on the model could be performed. However, the outcome can differ depending on the analysis goals, context, and individual perspective. In our experience, disagreeing with an assessment and justifying the opposing opinion already provides a deeper understanding. Therefore, we encourage reasoned disagreement with our evaluation.

4.1 Example 1: Surgical Training

The LAPSIM® by Surgical Science[2] is a commercial surgical training simulator for laparoscopic interventions. The system offers medical simulation training for effective and patient-safe training of surgical competence that can be transferred to the real operating room.[3] Laparoscopic surgery is a technique in which short, narrow tubes are inserted into the abdomen through small incisions. Long, narrow instruments are inserted and used to manipulate, cut, and sew tissue, as shown in Figure 5 (left). The LAPSIM is an advanced simulator with detailed graphics and haptic feedback to train these procedures. The system consists of a custom input device with two laparoscopic grips with precise tracking through a wire system and a third grip to control the camera position inside the virtual patient's body. The grips offer accurate haptic feedback for soft tissue and hard surfaces, such as bones, with force feedback delivered through the wire system. For the experience of being in a virtual operating room, the system is equipped with an Oculus Rift headset with outside-in head tracking. The Oculus controllers are not used. The virtual scene consists of a patient, a monitor with a view of the laparoscopic camera, and assisting surgery staff, e.g., nurses.

We now go through the loop clockwise, starting with the user's actions. The user interacts with the LAPSIM through custom laparoscopic grips. As a result, the user performs actions with their body that match well with real laparoscopic surgery, particularly the movements of the arms, fingers, and one hand. However, the other hand and the body posture differ as users stand comfortably in front of the LAPSIM while in the real surgery, they must lean in, as can be compared in Figure 5. Still, *action fidelity* can be considered high. The actions' detection is realized with a wire system precisely measuring the motions in four DoF (three rotational and insertion depth) of the laparoscopic grips—the same as in the reference. The system provides very high *detection fidelity* of the surgical tools. However, finger, arm, other body motions and the voice are not captured, decreasing *detection fidelity*. The system transfers the measurements into appropriate positions and orientations of the virtual grips. In addition, the system uses inverse kinematics to estimate hand and arm motions based on the end position of the grips. This gives the system high *transfer fidelity*.

Based on the input, the LAPSIM simulates tissue properties very accurately, such as its softness or response when cutting it. When blood vessels get damaged, the system simulates bleeding abstractly in the form of blood spilling out but not flowing anywhere. Outside the surgical site, the behavior of the surgical staff is simulated quite well as they perform relevant tasks in the operating room and react to the progress of the surgery. However, their animations seem sluggish and unrealistic. Therefore, *simulation fidelity* is in some aspects high, in others low. The VE in the LAPSIM can be considered as two separate parts: (1) the operating room the user is standing in and (2) the surgical site inside the virtual patient's body, which is displayed on a virtual screen within the operating room. The in-body view is rendered with highly detailed textures,

[2] https://surgicalscience.com/simulators/lapsim/
[3] Overview of the system https://www.youtube.com/watch?v=7wlIBBm1RXU

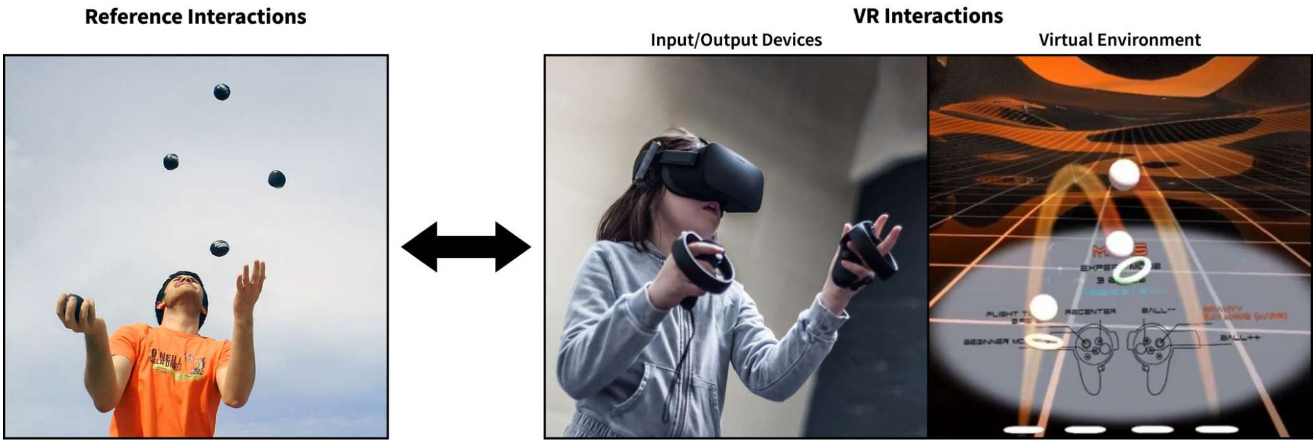


Figure 6: (left) The reference interaction of a person juggling with balls. (middle) A user juggling in the VR simulation using a VR headset and controllers. (right) The user's view within the VE: three balls are being thrown in a cascade pattern in a Mars-inspired environment. The two rings represent the user's hands. *Left image by Loris Bottello (CC BY-NC-SA 2.0, modified). Middle and right images (modified) from [4].*

reflections, and lighting for the internal tissue. On the other hand, the operating room and staff are rendered in a cartoon style that does not reflect the real environment well. Therefore, *rendering fidelity* is in parts high and medium.

The laparoscopic grips display the haptic feedback precisely by considering tissue softness and resistance. This aspect of *display fidelity* is high. In contrast, the visual and auditory output is displayed on the Oculus Rift headset, which differs from what is perceivable in the real world due to constraints such as a limited field of view or screen resolution. Olfactory cues are missing. Concerning modalities other than haptics, *display fidelity* is medium-low. Limited by each modality's display fidelity, the user's perception is equally restricted compared to reality. Hence, *perceptual fidelity* is in parts medium-low or high.

As a user's experience is highly subjective, we conducted an informal interview with an experienced surgeon to assess *experiential fidelity*. He has practiced with the LAPSIM and frequently conducts this type of surgery. The surgeon described the haptic feedback as extremely close to the real world, with a high contribution to the experience, as this is the focus of the intervention. On the other hand, the animations were described as not convincing. Some surgery procedures are missing in the LAPSIM, such as changing the physical instruments by completely pulling them out and inserting a new instrument. Overall, the surgeon assessed the system to have high realism as it comes close to the experience of a real laparoscopic surgery with a match of 70% to 80%.

In conclusion, the LAPSIM provides high fidelity for aspects that are most important for the training of hand-eye coordination, surgical procedures, and the development of manual dexterity. The system provides medium to low fidelity for less relevant aspects, such as staff animations and environmental graphics. However, this suffices for the training purpose, as scientific validation confirmed. Studies have shown that skills trained in LAPSIM can be successfully transferred to real surgery [18, 21]. Our detailed assessment of LAPSIM's interaction fidelity identifies the strengths and weaknesses of the system, confirms adequate prioritization in its development, and shows opportunities to improve realism further. The example demonstrates that thoughtful interaction and system design can help achieve the purpose of a VR system without improving fidelity in every aspect.

4.2 Example 2: Learning to Juggle

Juggling with three or more balls is a complex activity that can be challenging to learn. There is a steep learning curve as you either throw and catch the balls with the correct timing, or they will fall to the ground repeatedly. To make learning to juggle easier, a virtual simulation can deliberately deviate from the reference interaction in the real world, thus lowering interaction fidelity. In this example, we analyze the VR software *Planet Juggle* by Benjamin Outram,[4] which provides various features to facilitate a gentler learning process of juggling movements, such as the cascade pattern. For instance, the user can activate the following assistive features. Slow motion allows the user to practice and internalize movement sequences without getting hectic. When touching, the balls can snap to the hand to make catching easier. Visual indicators can show the trail of the balls and preview where they will go to achieve the ideal cascade trajectory. The balls can always reach the ideal height independent of the throwing impulse. And there is background music that helps get the ideal rhythm. When getting more confident, the user can turn off the features for closer-to-reality training.

Again, we now systematically look at the interactions' fidelity along the loop when all assistive features are activated. The 1-to-1 mapping of the motions within the 3D space would allow high correspondence of the body movements to the movements of real juggling. However, the grasping and releasing actions differ since the user controls the virtual balls with the trigger buttons of the hand-held Oculus controllers instead of physical balls. For

[4]Description, video trailer, and free download of *Planet Juggle* for the Oculus Quest and Rift at https://www.benjaminoutram.com/planet-juggle, last accessed 2023-08-15

holding the controllers comfortably, the hands are rotated inwards for virtual juggling, other than in real juggling, where the palms must face up to catch a ball. Further, supportive features such as snapping, slow motion, or automatic height make the system tolerant of discrepancies in hand position or rotation, throwing power, and bad timing. While this allows successful first steps in virtual juggling, the same movements would lead to failure in the real world, resulting in low *action fidelity*. The detection of hand movements works precisely using the Oculus tracking systems. Other body movements are not registered, although less relevant to the activity. Therefore, we can assess high *detection fidelity*. The system tolerates discrepancies in hand movements by making up for them with transfer functions. Although the user actions may differ from realistic juggling movements, the transformed virtual juggling actions closely correspond to the reference interaction, which leads to high *transfer fidelity*.

The lower gravity leading to the slow motion effect fits the narrative of *Planet Juggle* as the user is supposed to juggle on the Moon or Neptune. Disregarding side effects such as the dreadful death from cold, the simulation's physics is generally sophisticated regarding ball behavior. However, our reference interaction for this comparison is real juggling on Earth. Features such as slow motion or visual trajectory indicators limit realism. Particularly, the strong predetermination of where the balls can travel and land, such as the spatial confinement to a plane disregarding the third dimension away from the user, leads to low *simulation fidelity*. Moving on to output fidelity, the system renders an abstract environment with grids on the ground, geometrical landscapes, and glossy, highly reflective surfaces. Rhythmic music is generated to influence the juggling tempo, and the balls create abstract sounds when touching the hands. All this can be considered low *rendering fidelity*. The visual quality depends on the Oculus hardware used, but even with the superior Rift headset, the visual impression can be clearly distinguished from what the user could see in reality. Similarly, haptic fidelity is limited because the system only provides continuous contact forces from holding the controller and abstract vibrations when touching a ball. Overall, *display fidelity* is medium to low. The controller's vibration feedback acts as a sensory substitution. The user cannot feel the momentum and weight of the falling ball, but the vibration intensity is calculated from these parameters. The haptic cues can inform the user about the ball's properties and the impact of catching. Similarly, audible cues represent the ball's force on the user's hand. Consequently, *perceptual fidelity* can be considered higher than display fidelity, with medium fidelity.

Finally, to assess *experiential fidelity*, we tested the simulator ourselves with a think-aloud approach. The app gave a juggling novice the impression of quickly acquiring juggling skills as the movements for the cascade pattern were quickly performed. He felt competent to virtually juggle after a few minutes, thanks to the assistive features. On the other hand, transferring these skills to the real world was a completely different story, as too many aspects differ from juggling with real balls. Another of the authors, proficient in juggling, struggled with the skill transfer in the other direction. At first, he could not accomplish a stable juggling pattern because it felt so different from what he was used to. Mainly the haptics were found to be too different as the soft vibration gave no impression of an impact. Also, the lack of catching and releasing the balls with the hand felt unfamiliar. The author further missed the third dimension because the ball trajectory is restricted to a 2D plane. He disliked the assistive features in VR: "I find reality much more 'assistive.' I missed that in VR." Overall, the authors enjoyed the playful VR activity but did not get the feeling of real juggling. We assess *experiential fidelity* subjectively as low.

In summary, the simulator provides virtual aids to learn the principles of juggling by deliberately deviating from a highly realistic replication. The supportive features allow quick progress for virtual juggling and can help build muscle memory through authentic arm movements. A scientific evaluation of the app has explored how it can be integrated into learning to juggle with real balls [4]. The study showed how users had more fun training with the VR simulation but struggled with transferring the acquired skills. Therefore, developers should carefully consider which fidelity aspects should maintain high realism to ensure skill transferability.

4.3 Example 3: Lightsabers from *Star Wars*

In this third example, we turn to a reference interaction from a fictional narrative: using a lightsaber, the energy sword from the *Star Wars* franchise. Specifically, we examine the interaction fidelity of using lightsabers in the VR game *Vader Immortal: A Star Wars Series – Episode III* by ILMxLAB.[5] The player can use the weapon in lightsaber duels, to block blaster bolts, hit enemies, and throw it at targets. It is operated with hand-held Oculus Touch controllers. But what do we compare the VR interactions to if nobody has ever held a real lightsaber? The depiction in Star Wars media that also informed the game's development seems a reasonable match. Even in the best case, a fidelity assessment is debatable as we currently lack objective measures. Our evaluation is even more contestable when we compare to a fictional reference interaction since that already leaves room for speculation and disagreement. Interpretations can vary depending on the canonical choice of Star Wars media, such as comics, TV shows, books, video games, or merchandise artifacts. For this reason, we focus on the original movies as a reference. We invite all readers to question our stance and justify their proposal.

Instead of a circular procedure, we look at the most striking characteristics of the interaction. The weapon functions in the virtual world much as you would expect it to. The player holds the lightsaber's hilt, wields it, and throws it as the plasma blade blocks attacks and cuts through any objects. However, most objects are not cut into pieces when the blade goes through them, and it hardly leaves a trace. Due to the simplistic game mechanics of the choreographed lightsaber duels, also the opponents do not always respond to being touched by the lightsaber; similarly, the player does not die instantly from being hit by an enemy's lightsaber. In this game, using a lightsaber has medium *simulation fidelity*. The controllers as input devices are beneficial for *action fidelity* because the player has a similar hand pose as when gripping the lightweight, slim saber hilt. Even the activation by pressing a button corresponds well. How the player holds the weapon and performs combat movements with six DoF contributes to overall high *action fidelity*. It is sometimes lower, though. If the player holds the hilt with both hands, having two separate controllers limits

[5]Trailer and download at https://www.oculus.com/experiences/quest/2426206484098337/

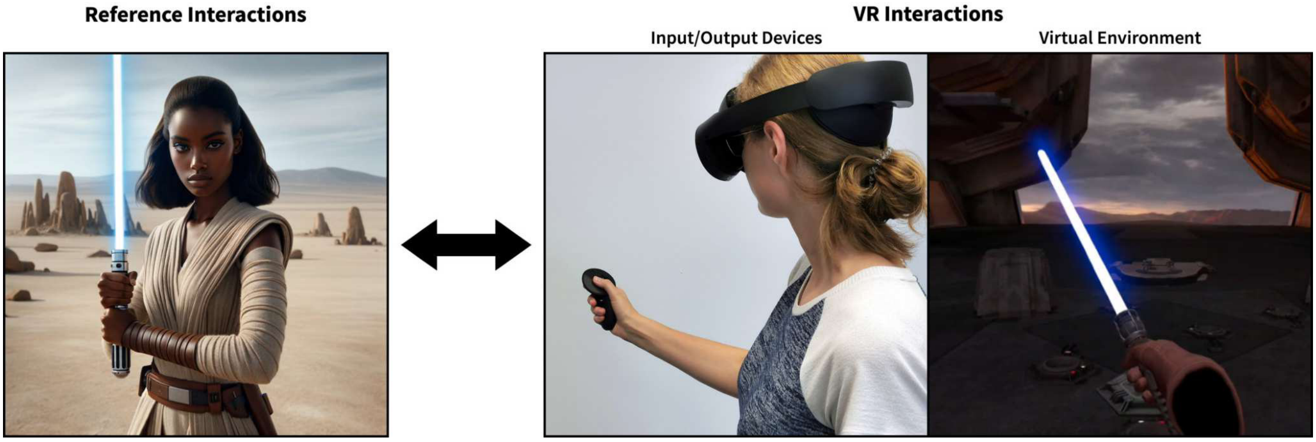


Figure 7: (left) A person holding a lightsaber as the reference to be simulated. (middle) A user with a headset and a controller using the virtual lightsaber. (right) The player's view in the VE. ***Left image generated with OpenAI DALL·E 3, right image captured from "Vader Immortal: A Star Wars Series – Episode III".***

action fidelity. It further suddenly decreases when there should be resistance from hitting another lightsaber, but the player's arm continues to move unhindered.

This equally affects *display fidelity* due to the lack of force feedback. Because the controllers only provide passive haptic feedback and vibrations, we assess this aspect as low-fidelity, even more so when virtually throwing the lightsaber but still gripping the controller. The headset's constraints on the visual impression reinforce this impression. At the same time, *transfer fidelity* can be considered high as the wrist's flicking motion and the grip button's release are interpreted as a throw, sending the weapon on a boomerang trajectory in an adequate direction. Generally, movements are recognized accurately, resulting in high *detection fidelity*. When activating the weapon, the player hears the iconic hissing sound. The pitch of the electric swoosh sound is higher when wielding the sword. Together with the convincing glowing effect of the blade, we thus assess *rendering fidelity* as high. Given the limited skills of the testing author, we are comforted by the low nociceptual fidelity as there is no pain from getting hit. Instead, the view is overlaid with a red vignette, and we hear the avatar moan, which could be considered a decreasing factor for rendering fidelity. Apart from that, *perceptual fidelity* is restricted by low display fidelity.

Lastly, to inform our assessment of *experiential fidelity*, we informally tested the game with a Star Wars enthusiast and someone who has only seen a few movies. The enthusiast was amazed at how much the lightsaber made him feel like a Jedi. It felt convincing, wielding a glowing, humming lightsaber in 3D space with high levels of embodiment, especially compared to playing the game series *Star Wars: Jedi Knight* on the computer with a mouse and keyboard or using sticks as a child. Only the crude combat game mechanics detracted from the experience. The other tester, who did not care about Star Wars, also enjoyed handling the lightsaber. When asked if this is how he would expect a lightsaber to feel and behave, he was unsure as he never contemplated it. He was surprised that there was no air resistance or weight like from a metal longsword but then assumed: "Probably, this is just what a lightsaber feels like." Furthermore, he was confused by the inconsistent controls: a blaster requires continued pressing of the grip buttons while the lightsaber only needs one short press. Also, the vibrations and the headset's visual limitations were described as irritating. Overall, both testers quickly forgot about the system's shortcomings and were immersed in the experience. We suggest a medium-high fidelity rating.

This example demonstrates that the reference interaction does not need to be based on the real world. As long as the original to be simulated is clearly defined, the IntFi Model can be applied meaningfully. Regarding experiential fidelity, users with no knowledge of the reference cannot make a competent comparison, but they still have their assumptions. Similarly, non-swimmers cannot compare virtual swimming to real experiences but can make an informed guess, assess the simulation's coherence and credibility, and form their impression intuitively. Moreover, the lightsaber interaction exemplifies how other factors (such as fun in gaming) should be prioritized over maximum fidelity. The game would hardly be playable if the weapon sliced every object and the self-avatar, when touching it with the deadly blade.

5 VALIDATION

To polish and validate the IntFi Model, we conducted expert interviews with 14 established VR specialists from academia and the industry. The goal of this evaluation was to collect feedback and criticism, discuss the proposed terminology, and put the model to the test in terms of its applicability.

5.1 Method

We conducted semi-structured, interactive interviews via Zoom or in person. We prepared an interview guideline to structure the conversations and make them comparable. The only variation between the interviews concerned the experts' example projects to which they applied the model during the interactive sessions. At least two of the authors were present for each session. We conducted the interviews between March and May 2023. The conversations took between 45 and 79 minutes, averaging 63 minutes. We recorded the sessions with the participants' consent and collected 14.75 hours of material. We performed a thematic analysis of the data to process and structure the findings. Two of the interviews could not be recorded due to a technical error, so we resorted to notes on these sessions. After finalizing the manuscript, we asked all interviewees whether the interviews had been adequately summarized in the reported findings. One expert did not find the time, but everybody else confirmed that the text accurately reflects the conversations without any requests for changes or additions.

5.1.1 Sample. We selected researchers, interaction designers, developers, and managers for their contributions to the VR field and their professional experience. We invited 19 selected candidates, of whom two could not find the time, and three did not respond. The diverse sample covered a broad range of VR-related research fields, backgrounds, and perspectives. Because we assured all participants that they would remain unidentifiable, we can only report broadly on the sample's composition.

The sample included five full professors, three assistant or associate professors, and a lecturer from academia, as well as two senior professionals, a research scientist manager, a product owner, and a consultant from the industry. Among the experts were distinguished scientists who were awarded three IEEE VGTC VR Technical Achievement Awards, the IEEE VGTC VR Significant New Researcher Award, an IEEE VGTC VR Best Dissertation Honorable Mention, and various career awards. The sample also included three IEEE VGTC VR Academy members and an ACM Distinguished Scientist. The ten researchers in our sample had an average citation count of 7,043 and an average h-index of 35 as of 14 February 2024 according to Google Scholar. The h-index ranged from 11 to 69 and reflects the spectrum of our selection that includes young scientists publishing with a focus close to the IntFi Model and experts with decades of experience in VR research. All participants primarily work in VR or HCI. The experts' focus includes haptics, locomotion, interaction techniques, embodiment, computer graphics, visualization, presence, training and education, avatars, perception, medicine, collaboration, artificial intelligence, and multimodal interfaces. The variety of topics demonstrates how broadly the model can be applied in practice. After analyzing the last interview, we invited three interviewees to join the author team and further contribute to this work, which they accepted.

5.1.2 Interview Structure. The interviews followed a semi-structured guideline based on a slide deck.

(1) **Working definition.** We defined the terms fidelity and realism in the context of our model.
(2) **Task 1.** Without any previous biasing, we asked the experts to describe the fidelity of the interactions in one of their works and with their own words. Find more details below on how we selected the work.
(3) **Introduction to the IntFi Model.** We explained the purpose of the model, the objective approach, the fidelity spectrum, the HCI loop as the foundation, and the components of the model integrated into the loop. We then elaborated on the single aspects with definitions and characteristics.
(4) **Initial feedback.** We asked for first thoughts, any confusion or questions, criticism, and concerns regarding the terminology. In some conversations, this feedback was already raised during the presentation of the model.
(5) **Task 2.** We suggested different ways to apply the model. Once more, we asked the expert to elaborate on the interaction fidelity of the same work as in task 1, only this time using the IntFi Model as a basis. We discussed it in more detail and asked follow-up questions as needed, such as identifying key aspects with prioritized fidelity, low-hanging fruit for improving fidelity, connections between aspects, etc.
(6) **Conclusion.** Finally, we asked for the expert's overall evaluation of the model. Depending on the time left, we went into more detail regarding possible benefits, improvements, or research opportunities that the experts wanted to add.

For the two tasks of assessing interaction fidelity in a specific example, we selected one project, system, or user study of the expert. The interview partners were familiar with the work, so we could directly dive into the discussion. This approach also allowed us to authentically test the practical applicability of the IntFi Model in genuine use cases. By using the example in both tasks, we could compare how the experts approached the analysis either with or without the model providing structure. While we observed the differing levels of detail and evaluation strategies between the first and second task execution, the experts could experience how the model can facilitate the assessment of interaction fidelity.

The works covered a wide variety of topics, including avatars, locomotion, haptics, perceptual illusions, presence, object manipulation, social interactions, embodiment, emotions, virtual environments, and training. To preserve the anonymity of our interview partners, we do not specify the chosen publications or projects. The works were required to link to the model, i.e., depend on at least two aspects of fidelity, aim for any kind of fidelity, or be inspired by an explicit reference interaction that has been reproduced.

5.2 Findings

In this section, we present the insights from the expert interviews. We refer to the experts from our sample with E01 to E14. Overall, the interviewed experts found the IntFi Model "useful" (E04), "comprehensible" (E07), "sound" (E02), and "a meaningful contribution" (E03). However, there was also reasoned criticism, opposing views, and a need for clarifying discussions. For example, some terms were described as "not perfect" (E11). Expert E10 was not convinced of our validation process' rigor, while E04 particularly liked it and found it "very systematic". E08 struggled with the HCI loop as a not entirely suitable basis but also found that "some more modern reflection on the old interaction loop is definitely interesting." Each of our interview partners appreciated the IntFi Model as interesting and helpful.

The findings are structured by the identified themes: *concept criticism*, *terminology criticism*, *applicability criticism*, *application strategies*, and *contributions*. The themes *application traps* and *patterns* were moved to the separate Sections 6.3 and 7.1. Some critiques are addressed in the discussion section and, therefore, only mentioned here as a topic but not directly elaborated upon to reduce redundancy. Comments, misunderstandings, and concerns that were resolved in the interviews are not outlined. Similarly, any additional literature suggested by the experts was integrated into the Related Work Section but is not mentioned here separately. We directly adjusted the model and this paper according to the experts' suggestions without reporting before-and-after comparisons to avoid confusing readers with changes from an unknown earlier version. These unreported improvements primarily concern details of descriptions, definitions, framing, or illustrations.

5.2.1 Theme: Concept Criticism. The fundamental conceptualization of the IntFi Model, linking the fidelity aspects to the stages of the HCI loop, convinced most of our interview participants. E03 found that "on a conceptual level, this is a very nice piece of work. I think it would be a meaningful contribution to how we think about and talk about virtual reality systems. [...] It's obvious a lot of thought went into this." One participant called the model "a very high-dimensional matrix" (E08).

However, there was also criticism on a conceptual level. The two most prominent critiques addressed the claim of objectifiability and the lack of quantification. Concerning the former, some participants doubted that all aspects of the model can be described objectively and struggled with the impracticality of ever determining an underlying ground truth. E14 hypothesized that system-related aspects (i.e., the bottom five) can only be assessed objectively, while the user-related aspects (i.e., the top three) only subjectively. We argue in the discussion (Section 7) that all components can be assessed objectively and subjectively depending on the approach. Concerning the latter, most experts indulged at some point in the interview in the idea of how helpful it would be to quantify interaction fidelity and its components: "This scale from low to maximum fidelity: this is a scale from 0 to 100, and I want to have that value. It would be nice if I could assess that, of course." (E14). Enjoying the prospect of universal fidelity metrics and measurements ourselves, we explain in the discussion why this is not in the scope of this paper and might need many more years of systematic research.

For some interviewees, the concept of fidelity was less intuitive than realism. They initially proposed reframing the model for comparisons to real-life interactions as the most common replication reference of VR simulations. After looking at practical examples with reference frames other than reality, most experts supported using the fidelity concept and preferred the more universal applicability. One of the FIFA authors criticized this limitation in their own work in hindsight: "That was a weakness we had with [FIFA] that we were gauging everything with regards to the real world."

One expert from our sample, E08, raised doubts that it makes sense to consider perceptual fidelity "because we can't know—it's unknowable!" Although it is currently challenging or impossible in many cases to determine what a user objectively perceives, there is a ground truth to it that has a higher or lower match to what the user would perceive in the reference interaction. With this, independent of its ascertainability, perceptual fidelity is a valid construct. Some interviewees pointed out that perception might even be measurable in the future if we develop more sophisticated neural interfaces. E08 continued, that "obviously, nobody knows if perceptions are the same. That's the whole philosophical debate about qualia." For this reason, we recommend considering perceptual and experiential fidelity on an individual level. Only a user-by-user assessment can do justice to different personal abilities and characteristics. From a practical stance, E05 argued that while perceptual fidelity "is user-dependent, it is not unknowable or unpredictable if you have enough information about the user."

Another critique by E08 concerned the HCI loop as it is segmented into stages, while interactions in VR often occur with simultaneous input and output in parallel. Appreciating the debate about the loop's limited applicability for direct manipulation techniques due to the concurrence of input and output, we argue that it is nonetheless insightful to use the loop as a theoretical construct for abstracting and distinguishing the involved elements of the reciprocal VR interactions. Breaking down the permanent exchange of user and system can help understand the ongoing parallel processes while identifying isolated aspects relevant for analysis.

Fundamentally, E08 was skeptical of operationalizing fidelity at all "as it's obviously impossible to reproduce things." In our conversation, we agreed that perfectly replicating an original is extremely difficult and potentially impossible, especially regarding something as complex as bodily interaction with the world. However, using the IntFi Model to determine interaction fidelity is most helpful on the vast spectrum before reaching "maximum", i.e., to describe *how* imperfectly something is reproduced. Realistically, we are nowhere close to the upper extreme of the spectrum with current technology. Until we achieve perfection, the model can help assess the degree of fidelity.

Several experts wondered how some aspects can be evaluated independently of others as they seem inseparably linked in practice. For example, rendering fidelity seems to be coupled to simulation fidelity on a computational level and to display fidelity on a hardware level. We agree that from the optimization viewpoint, the components should strongly depend on each other. Consequently, they are usually configured in combination by game engines and developers. Still, it can be beneficial to consider the aspects individually from an interaction design perspective. Consider a use case where a user cannot see the hair on his arm. It makes a difference if it is incorrectly represented in his avatar model, if it is not rendered due to missing height maps, or if the display resolution is too low for the hair to be recognizable. Further, distinguishing the aspects can even be required from a technical point of view. For example, in cloud computing or when watching 360° videos, the rendered output can be displayed on HMDs with different screen resolutions, affecting rendering and display fidelity separately.

As an addition to the model, E03 proposed integrating a ninth component along the middle axis of the loop: an element that moderates between the user's mental model (as part of experiential fidelity) and the system's data model (as part of simulation fidelity). We have not implemented the proposed extension of the IntFi Model for two reasons. First, there is no correspondence of the abstract concept of a mental model to any reference interaction, which is why the term "fidelity" does not apply. Second, the HCI loop has no equivalent component, and we prefer sticking to the original conceptual foundation.

5.2.2 Theme: Terminology Criticism. We asked the experts about the clarity and suitability of the chosen terms. Most interviewees had no concerns regarding any of the labels and appreciated how consistent they are. E02 said: “I think the terms are perfectly clear as they are. And it’s good that they’re short, both in writing and when talking about it.” Although many experts pointed out how introducing a defined vocabulary can be helpful for communication, others cautioned against conflicts with existing terms or definitions. In the literature and oral vocabulary, some terms are used differently from the IntFi Model understanding, and sometimes different terms are used for the same thing. For example, *interaction fidelity* is coined more narrowly in the Framework of Interaction Fidelity Analysis [59]. One of its co-authors, who is part of our sample, commented on the redefinition: “In our previous work, we used [interaction fidelity] almost exclusively for the input side. But it was always a problem. [...] That will be a battle you’ll have to fight to get the term to mean more. But it makes sense overall.”

Furthermore, there was criticism on a few specific terms, mainly by E11. We thoroughly discussed the critique with the experts and among the authors. In the end, we chose the term that seemed most suitable. For example, we changed *sensory* to *perception fidelity* to avoid confusing it with system sensors, and later from *perception* to *perceptual fidelity* for linguistic reasons and consistency with *experiential fidelity*. We also changed the previous term *feedback fidelity* to *output fidelity* for higher symmetry with input fidelity and a better match with the component *output devices* of the HCI loop. Moreover, the term *feedback* implies a reaction of the system to user input. Hence, in a non-interactive 360° movie without user input, there can be no feedback, but the system can still give output.

Other alternative labels were suggested in the interviews but not adopted after careful consideration. For example, *tracking fidelity* was suggested instead of *detection fidelity*, though this term implies including the transfer component. Alternatively, it was proposed to be called *human sensory fidelity* while *technical sensory fidelity* would replace *perceptual fidelity*. However, this might facilitate misunderstandings and would be less concise. Further, the label *action fidelity* was criticized for not encompassing passive states and uncontrolled behavior by the user. Instead, *behavior fidelity* was suggested, although this has the same limitations. Alternatively, *body fidelity* was suggested. However, this has an overly passive framing and could be misleading by implying only to represent the user’s body as an avatar and neglecting voice or neural input. None of the terms are perfect, so we chose *action fidelity* as the most intuitive and flexible compromise.

The experts also criticized terms for which we could not find any better alternative. For instance, *input fidelity* was challenged because undetected actions never enter the system and, thus, are not indeed system input. We argue that the user’s intent and the potential of capturing all actions are decisive for being included in *input fidelity* as the user would expect a high-fidelity system to detect everything. Lastly, we observed that *display* and *rendering fidelity* have strong connotations with visual screens for some people. In our sample, this was especially pronounced in the industry. In the instructional material, we emphasized that any modality can be rendered and displayed.

5.2.3 Theme: Application Criticism. With an example of their own previous work, all experts utilized the IntFi Model to analyze and reflect on a user study or system design. This enabled them to assess its applicability in a realistic scenario. Some of the experts experienced the onboarding as demanding: “It’s a lot to take in. [...] If I would have to use it as a tool, I would require a bit of time to get familiar with all the intricacies” (E02). It was a challenge for the participants to understand the model’s concept and remember the components’ definitions. Some of them grasped the ideas quickly as they studied the subject intensely. To others, the proposed way to contemplate fidelity was unfamiliar, so they needed more time, detailed explanations, and examples to wrap their head around the model. Most participants asked several comprehension questions during or after the presentation of the model.

Even interviewees experienced in fidelity research recommended creating accessible instruction material: “ideally, a graphical version that would make it very easy to use” (E02). To enable teams to use IntFi Model as a practical tool, there must be a quick and easy way to learn how to work with it. Therefore, we developed an informative and intuitive poster that gives a visually appealing introduction to the model. E02 suggested “creating a very nice graphical representation of it that you can print out in A0 and have in your lab. [...] Having all of it up on a wall would be very good for me.” This agrees with our experience of hanging it on the wall as a poster and referring to it during our daily work. We also prepared a slide deck that can be conveniently adapted for teaching material. Both can be accessed in the appendix and used for free under a Creative Commons license.

All experts but two stated that they would like to use the model in their work in the future. One of the two, E08, wanted to read the paper first. The other expert, E10, was concerned about adequate validation before employing it: “I think validation is important when it comes to something like this.” E04 appreciated the expert interviews as a validation method: “I like the approach. Very systematic.” For E10, however, to sufficiently verify the soundness and acceptance of the proposed model, it must prove itself in practice on a large scale. Only if the community can work with it and embraces it the IntFi Model will be valid to E10. At the same time, this seemed to be the most promising prospect: “At the moment, it’s a theoretical model, a conceptual model. Making it more than that, that’s a research opportunity. Bring it to practice!”

When applying the model, some experts fell into application traps, which led to criticism of the model’s applicability. We collected these traps in Section 6.3 as guidance for readers without elaborating on them further here. We also collected special use cases and integrated them at appropriate parts of this publication, such as the applicability for multi-user systems (E06), mixed reality (E13, E06), or troubleshooting (E12, E14).

5.2.4 Theme: Application Strategies. From the observations in our interview workshops, we can conclude that the IntFi Model successfully guided the participants through the analysis process and was a helpful basis for discussion. This is evident in the different approaches and outcomes from the first and second tasks, i.e., without or with the help of the model. In the first task, without the model, the participants’ explanations and considerations were unstructured, and occasionally seemed lost in the complexity of the subject. Most were uncertain about where to start and stopped after mentioning only a few aspects. In contrast, equipped with the model in the second task, all participants identified many more relevant aspects and kept elaborating on the interactions' fidelity. Typically, it took them a few moments to think about it and match the model with the use case. Then, they went through the single aspects and their connections, talking like a waterfall and coming up with additional insights they had not considered in the first task. Further, they now involved every part of the interaction in their analysis. E12 described this experience with: "I thought it was really cool because [with the model] you could just work through several points step by step, whereas before I just groped in the dark." With this, the IntFi Model seemed to have an empowering effect on our participants.

The experts used different strategies for applying the model to their individual use cases. We did not instruct them how to proceed but just set the task to describe interaction fidelity using the model. Some explicitly asked for the correct way to use it: "Where does it start? On the top with the user? Or with action fidelity?" (E12). Indeed, half of the sample adapted the approach we chose for presenting the model, starting with action fidelity and continuing in the loop aspect by aspect. As a variation, E02 started with action fidelity but then followed the logic of the interaction technique. Alternatively, three experts started with their study's independent variable and went in the loop from there. Similarly, E01 started with the aspect most important for the application but then went through aspects chaotically and went back and forth, addressing different modalities. E04 and E09 had a purely chaotic approach, following their narrative organically as it evolved. E14 deliberately made two rounds in the loop: first for a "clean" version of the interaction technique, then for the modified version with sensory manipulation. This helped him contrast the differences.

It was our impression that the experts from the industry generally needed more information and assistance for applying the IntFi Model. For example, E12 and E07 asked to return to the introductory slides with the descriptions for performing the second task. As a result, their analysis was conspicuously close to the notes and talking points on the slides. Although it seemed more demanding, they appreciated how beneficial it was. E09 concluded that "a more analytical approach for designing [immersive] interfaces and experiences would be reasonable because I think a lot of designers shoot in the dark."

The experts in our sample used the model for different goals. In most cases, they roughly located all aspects within the fidelity spectrum as a starting point. After this analysis phase, they interpreted their findings depending on their goal. For example, some attempted to reason or predict experiential fidelity through the other aspects (E01, E05, E06, E09, E12). Others wanted to set informed design priorities by weighing the aspects regarding their relevance for a use case (E01, E06). For this, they also linked their interpretations to established design principles and effects, such as visual dominance. For some researchers, assessing the degree of fidelity was less decisive. Instead, they wanted to identify which components the independent variables of a user study were part of (E03, E14): "This is helpful! Because it makes more precise what I did and did not manipulate in my research" (E03). This researcher realized that all of their manipulated study variables affect simulation fidelity, which they first criticized as not that interesting. Still, while contemplating using the IntFi Model, they realized they had achieved what they set out to do in the study: alter aspects of the interaction influencing plausibility illusion. Since E03 interpreted this as the relationship between simulation and experiential fidelity, they concluded to have chosen a suitable focus. In a further step, some researchers reflected on aspects that would be interesting to manipulate in follow-up studies (E02, E06): "It has been very useful trying to understand what it is I'm manipulating, and when designing an experiment, figuring out what are the factors that are relevant to manipulate" (E02).

A further reoccurring strategy of applying the model was identifying aspects of utmost importance for specific use cases or purposes (E06, E09, E13). For example, in applications for motor skill training, it is critical to optimize for high action fidelity, while simulation fidelity must be as high as possible for educational purposes. Therefore, some participants tried to define target variables, such as learning success, confidence in use, fun, or control precision, to deduce the determining components. Overall, the strategies and paths of applying the IntFi Model were as varied as the participants' objectives.

5.2.5 *Theme: Contributions.* Although the purpose of the interviews was to obtain criticism from experts in the field, we also want to outline the commendatory remarks of our participants. This theme includes positive remarks, suggested use cases, and contributions of the IntFi Model brought up by the experts.

Overall, the interviewees appreciated the model: "It's a nice theoretical framework. It's well polished" (E04). Many of them enjoyed contemplating and discussing VR interactions with the model as the basis of the conversation: "Interesting! That gives me some thoughts" (E08). The model was described as "super cool! super exciting! super useful!" (E06), and E07 concluded that it "makes sense! It's all comprehensible". An interviewee from the industry liked the IntFi Model as it helped reflect on fidelity in VR: "It's interesting because I understand it intuitively, but I don't know how to formalize it. And this is spelling it out for me. [...] I love it!" (E09).

E02 acknowledged that the model builds on existing frameworks and concepts: "I like the model. I used similar models myself, but this one takes it a step further and adds detail. I think it makes really good sense. [...] It adds structure to a large body of literature. It distills a lot of different concepts and presents them in a single model. The fact that we can go one level deeper and get more subcomponents, I think, has a lot of utility—both for designing studies and as a pedagogical tool." Contrasting it to similar frameworks, E11 liked the "agnostic approach". Extending previous work, E05 found that "comprehensiveness and consistency around the entire loop of VR interactions is the main benefit." In particular, the advancements compared to the Framework for Interaction Fidelity Analysis (FIFA) [59] were addressed in the interviews. While the FIFA only considered input, the IntFi Model also considers the system output. Also, FIFA's focus on the system was extended to the user and their mutual influence. And while FIFA was limited to realism, the IntFi Model extends to any reference interaction, from reality or not.

Further, it seemed important to many of the experts "to have a precise and integrated set of terminology. I think that's good and the biggest benefit" (E03). For unambiguous communication within the community, participants found it "very useful to get a shared vocabulary" (E02). Beyond using the same labels, E13 emphasized: "The benefit is not only to talk about the same thing but also to trust the other person refer to the same thing." While we also received critical feedback and alternative suggestions, most participants liked the chosen wording. The terminology was described as "accurate" (E03) and "consistent across the model" (E02).

Most of the experts were eager to use the IntFi Model in their work, described it as "helpful" (E01), and found various use cases: "I like this a lot in many ways!" (E05). Many of the proposed applications concerned system design and development. One utility often mentioned was the help in understanding the single components of interaction fidelity to predict experiential fidelity or the user experience. Distinguishing the fidelity aspects was also described as helpful in setting priorities depending on the goals or purpose of a system. As a next step, the model was considered helpful for iterating system designs to increase interaction fidelity. Also, participants liked comparing two similar systems based on the model.

To analyze existing systems, the IntFi Model was used by some participants like a checklist: "It's cool to think it through in small steps: like check boxes that you can tick off" (E12). Also, E01 emphasized: "The structure is helpful when going through the individual fidelity categories." Similarly, troubleshooting was mentioned as a use case. If there is an issue with the system or the users are unsatisfied, E12 and E14 suggested using the model to search for the problem systematically: "I like how you can examine a user's experience in smaller steps: Where exactly now does the error get into the system?" (E12). E06 and E14 further suggested use cases outside VR, such as for mixed reality interactions, video instructions, or other applications with a simulation approach.

Furthermore, the experts suggested numerous ways the model can be used for research. In the interviews, many researchers used it to understand a study better in hindsight: "What is it that we manipulated in our study?" (E02). We observed many instances where the experts tried to identify the modified independent variable of a study and how they searched for dependencies within the loop. This was also considered helpful for planning upcoming studies. Several researchers suggested a systematic literature review to identify which fidelity aspects were investigated in VR user studies. The model could provide a structure for a large body of literature (E02), and it could be expanded to be a signpost to all related work with details on specialized subcomponents (E13). Nine interviewees mentioned the significant research opportunity of systematically studying the fidelity components across the loop, both individually and regarding their influence on each other. Due to the model's structure, high comparability might help identify patterns (E05, E06, E13, E14), such as uncanny valleys in different sensory modalities (E04, E12). These endeavors were often considered a long-term community effort.

Lastly, teaching was mentioned repeatedly as an ideal use case for the IntFi Model: "Excellent presentation! I would include this in my curriculum right away. These are good learning materials for the students" (E13). The participants liked how they could convey the processes and connections of different factors in VR interactions with one central figure as an overview.

6 APPLICATION OF THE MODEL

In this section, we illustrate ways to use the IntFi Model in the work of researchers, designers, developers, practitioners, teachers, and students who work with VR. We collected best practices from our own experience and the expert interviews (see Section 5), possibilities for applying the model as a tool, and common traps to be careful of.

6.1 Best Practices

We recommend the following guidelines for applying the IntFi Model purposefully and effectively. They build on experiences from using the model for over two years and observing the participants in our expert interviews use it.

- **Mind your reference.** It is crucial to keep the reference interaction in mind when assessing fidelity. Define the reference as specific and detailed as possible before comparing it to the VR interaction. Keep recalling the reference interaction throughout the analysis, not only at the start. When using several references simultaneously, be aware of which one you currently compare to.
- **Choose a focus.** Decide how holistic or focused your analysis should be. Are you interested in just one isolated aspect of the experience (e.g., a handshake) or all elements involved (e.g., the full complexity of greeting conventions, environmental conditions, multiple users, etc.)? It can be easier to break down the interactions for analysis.
- **Set a goal.** Reflect on the purpose of applying the model. The IntFi Model can be used as various lenses through which interactions can be examined, allowing different perspectives. We propose a number of ways to apply the model in the next section.
- **Skip irrelevant aspects.** You can use the IntFi Model modularly. In most cases, only parts of the loop are needed for analyzing an interaction. Feel free to ignore irrelevant components and instead focus on the key aspects.
- **Be objective.** The connotation of "high-fidelity" as "better" is common in practice as the *Better–Worse Trap* in Section 6.3 illustrates. However, the IntFi Model works best as an objective tool. The fidelity concept is free of judgment. Therefore, clearly differentiate between "high experiential fidelity" and "great user experience." While higher fidelity can be desirable for interactions, deliberately decreasing fidelity can also help reach a goal.
- **Justify your assessment.** Be more specific than only assigning low- or high-fidelity labels. The insights from analyses or discussions can be richer if you argue how you came to that conclusion. This can help identify dependencies or patterns.
- **Adhere to the terminology.** Please stick to the official terms used in the IntFi Model when referring to it in scientific communication. If you need to specify a fidelity subcomponent (e.g., haptic fidelity), additionally mention the higher-level component it belongs to in the model (in this example, display fidelity).

6.2 Lenses of the Model

The model can serve as different lenses through which interactions can be viewed. Depending on your goals, the IntFi Model can be applied with various strategies yielding different insights.

Describe & Report: The IntFi Model provides well-founded and consistent terminology for the different aspects of fidelity in VEs that all contribute to the overall fidelity of a system. This structure helps to identify and name the fidelity aspect of interest and provides a foundation for an informed and differentiated discussion about fidelity in VR. In communicating study results, the investigated realism dimension can be unambiguously specified, supporting comprehension and retrieval by other researchers. Furthermore, the model offers a structure for reporting VR system specifications rigorously within the VR research community.

Understand & Distinguish: The IntFi Model can be used to understand the influencing factors on the overall fidelity in VEs at different stages of the HCI loop. It facilitates a comprehensive understanding of distinct factors and typical relations between the single components that contribute to the overall fidelity in VR. The model can also help to untangle interwoven components such as input and output through the same device. For example, user actions and haptic feedback when using force-feedback gloves directly depend on each other. Designers and developers can apply the framework to understand why their product performs the way it does and make informed decisions that explicitly address the influences of the single fidelity components on the overall fidelity. This could, for example, include a coherent level of fidelity across all stages of the loop or strategies to compensate for limitations in one component through improvements in others, e.g., when bodily trembling in an interaction can be compensated through denoising the signal with the transfer function. Overall, educators can use the model to explain the complexity of fidelity in VR and give students an overview of approaches for the different components of fidelity and examples of how these typically unfold in combination.

Compare & Analyze: Researchers, designers, and developers can apply the IntFi Model to compare variants in VR setups, e.g., different input devices, and systematically analyze the effects on the distinct fidelity components. As the model holistically addresses input fidelity and output fidelity, a complete reflection of the effect of, e.g., joystick vs. actual walking as different input techniques for locomotion in VR not only addresses the fidelity of the users' actions, the detection and the transfer function on the input side. It also directs to connected aspects of the output, such as perceptual fidelity, which can broadly vary between setups with different input fidelity. The model's components can provide insight into where differences or similarities between systems are, how decisive they are, what the underlying reasons are, or what could be done to compensate for them.

Hypothesize & Guide: The IntFi Model can be used to generate research questions around fidelity in VR to analyze and understand this design space further. It reveals many research opportunities, such as design guidelines regarding the combination of different fidelity levels in the model's components. For example, *can components compensate for each other? Or should the components rather have coherent levels of fidelity? How do the quantitative components influence the overall experiential fidelity?* The model could also be used as a starting point for heuristic evaluations of the fidelity of VR systems by systematically addressing the fidelity components. For empirical evaluations of realism in VR, researchers can also use the IntFi Model to formulate reasoned hypotheses as well as to explain and discuss the findings in relation to the different stages of the loop. In the following section, we will elaborate on the arising research opportunities based on our model.

Teach & Convince The model's simple structure combined with the intelligible visualization makes it easy for students to learn about VR interactions and how different components must be considered for reproducing something virtually. As the IntFi Model can be universally employed for any use case, it works in various practical and scientific curricula. Teachers can use the sequence of aspects to guide students stage-wise through the relevant components of interactions while emphasizing the reciprocal nature of human-centered simulations. Similarly, the model is suitable for convincing stakeholders and managers with limited experience with HCI methods or VR technology why a proposed strategy, system design, or research agenda would be advisable. The complexity of seemingly trivial interactions can be demonstrated just as effectively as the interdependence of the single components.

6.3 Application Traps

We identified several common pitfalls that people applying the model fall into and which we were also repeatedly caught in when developing it. The following traps are partially based on experiences from the expert interviews presented in Section 5. Be sure to avoid these common mistakes to get the most out of the IntFi Model.

Better–Worse Trap The most frequent fallacy might be attributing a judgment instead of objectively describing fidelity. People tend to use phrases such as better, worse, nicer, more immersive, better UX, etc., instead of an impartial assessment of the exactness of correspondence. The neutral, dispassionate view in fidelity evaluations is not the most intuitive attitude. A possible explanation is that fidelity and desirability correlate for many interactions, especially when striving for natural interfaces. However, it can also be beneficial to decrease fidelity deliberately to achieve a particular effect. Therefore, it helps to avoid thinking of low fidelity as a shortcoming.

Time-Travel Trap In contrast to absolute assessments of how close an aspect is compared to the reference, peoples' assessment is sometimes linked to the state of the art at a certain time. For example, some researchers evaluated a system in the context of technical possibilities at the time of development. Consequently, the low-poly visuals of an application were described as high-fidelity because 15 years ago, when it was built, the system was considered world-class, thus high-fidelity compared to anything else at the time. It can be reasonable to make a time-dependent comparison, e.g., when selecting hardware with the current technical limitations or tracing system capabilities over time. In other instances, however, it can be misleading to make assessments depending on the current state of the art as it shifts the assumed upper limit from *maximum fidelity* to *currently attainable fidelity*.

Apples-and-Oranges Trap Another trap we experienced and observed frequently is comparing the VR interactions to different references without noticing. For example, in the expert interviews, one participant assessed a locomotion technique as low- and high-fidelity at the same time and was confused about the outcome. The reason was that the comparison was made to different references: first to a teleportation technique, second to actual walking. While both comparisons can result in valuable insights, it should always be clear which reference is currently compared to. Another person fell into the same trap and referred to it as "comparing apples and oranges." Therefore, it is vital to apply the IntFi Model in a meaningful way. When comparing, choose a clear purpose and an adequate reference interaction. When in doubt, you can ask: *What* do I want to reproduce?

Visual-Dominance Trap As there is a strong connotation of *rendering* and *display* regarding vision, there is a risk of neglecting other modalities and senses when using the model. In many experiences, the visual impression is the most dominant and sophisticated modality. But it might be important to include all senses, depending on the use case. On the other hand, it can also make sense to narrow the focus to relevant modalities.

Feasibility Trap We also observed the risk of overestimating a system's fidelity and referring to it incorrectly as *maximum fidelity*, although it was merely the highest fidelity possible to achieve. Increasing it further could be restricted by limited resources, physical boundaries, personal abilities, or one's own imagination. However, it is irrelevant to assessing fidelity if there are limits in feasibility. Describing the exactness of correspondence is technology-agnostic, hence independent of the reasons behind a system design. To achieve maximum fidelity, optimizing a system as much as possible is insufficient unless there is a perfect match with the reference. We may never accomplish maximum fidelity in some aspects.

7 DISCUSSION

We now turn to the broader context of the model, considering its general meaning and implications. In the following, we discuss potential patterns of how the fidelity components might be connected, considerations of optimizing for realism, how the IntFi Model relates to similar constructs, how fidelity can be described objectively, how it can be measured, the significance of perceived fidelity, and how the model can be applied to reference frames other than reality, such as fiction or mixed reality.

7.1 Patterns

Investigating numerous VR interaction techniques, systems, experiences, and user studies, we found similar phenomena repeatedly. From this, we distilled reoccurring connections, dependencies, and relationships between the fidelity components. In this section, we propose potential patterns that might be discoverable in various VR interfaces. Not all patterns necessarily appear in every interaction. Please note that the patterns presented here are not systematically studied and lack empirical evidence. They are merely based on incomplete sets of examples and theoretical reasoning. Further systematic research is needed to test these suggested patterns and reveal additional ones.

Bottleneck Pattern | In some use cases, experiential fidelity cannot be higher than a limiting key component, even if other aspects have much higher fidelity. For example, when juggling in VR, it is irrelevant if visual rendering fidelity is exceptionally high as long as action or simulation fidelity is low. The bottleneck of these limitations will always impair experiential fidelity. Consequently, it is necessary to identify the critical components of an application and prioritize them for increasing fidelity.

Loss-Propagation Pattern | Certain aspects in the loop cannot be higher than the previous one. In this case, the component inherits the constraints of the preceding one. We can find such a conditional dependence in the pipeline *Simulation → Rendering → Display*. For example, a display can depict something at most with the resolution it was rendered at, and the rendering software can, at best, match the quality of the simulation's 3D model. Although we can speculate and approximate to compensate, we cannot assure the compensation's authenticity. Thus, loss at an early stage of such sequences cannot reliably be made up for and is propagated throughout the loop.

Similar dependencies can be in the sequence *Action → Detection → Transfer*. For example, if a controller prevents a natural hand pose when juggling, no sensor of an input device can make up for this, and if the sensors do not detect the hand position, a transfer function cannot compensate for missing data. There might be further such sequences, e.g., *Display → Perceptual → Experiential*; or *Experiential → Action*. As a consequence, it is advisable to optimize at the start of such sequences to avoid inheritance of early losses. However, a component can also be limited by a component much earlier in the loop, e.g., *Action → Perceptual* due to the vestibular system. Therefore, searching for the root of an issue in the preceding components can be helpful, as it might just be a propagated problem.

Irrevocable-Loss Pattern | An extreme version of the Loss-Propagation Pattern was proposed by E08 in our expert interviews: "It's like a pipeline. [Progressing through the pipeline], there are only losses." In many use cases, such a drastic error progression might occur. However, it can be prevented in some cases. For example, if the hand pose is detected incompletely due to occlusion, we can still compensate for this deficiency by reconstructing a probable hand pose using anatomic models and inverse kinematics. Consequently, there might be a viable solution to a component's limitation later in the pipeline, compensating for earlier losses.

Free-Upgrade Pattern | Maximum fidelity in some components automatically leads to maximum fidelity for the subsequent component. For example, if the displays reproduce the physical stimuli perfectly, they will also be perceived indistinguishably from the original stimuli of the reference, as we cannot sense the source of a physical signal but just the signal itself. Thus, maximum display fidelity inevitably results in maximum perceptual fidelity. Similarly, experiential fidelity will automatically be at maximum as inherited from the previous component. Therefore, if you need to increase the fidelity of a particular component, it might be helpful to optimize the preceding components.

Uncanny-Valley Pattern | Comparing varying input fidelity with user performance, McMahan et al. [59] suspected an "uncanny valley of VR interactions" that leads to poor performance for unfamiliar interfaces with medium input fidelity. While the tested low-fidelity interfaces were known from preexisting systems and the high-fidelity interfaces were intuitive, both delivered higher performance than unknown, somewhat abstract medium-fidelity alternatives. However, as the authors discuss, this non-linear relation cannot be generally applied to any input system. Further studies identified similar patterns in gaming [50], training [12], and locomotion [65]. Consequently, it might be beneficial to avoid interaction techniques with medium fidelity.

7.2 Optimizing Realism

Reality is not the only reference frame to which the IntFi Model helps compare VR, but it is undoubtedly a particularly relevant and common one. Therefore, we discuss the ambivalent goal of striving for high realism. An interaction's level of realism is merely a descriptive, impartial attribute. High realism is not *per se* superior. For some use cases, increasing interaction realism can be beneficial (e.g., skill training and education) or fundamental (e.g., preservation of historical artifacts). For other use cases, enhancing realism might be negligible (e.g., data visualization, art), detrimental (e.g., fictional entertainment), or even harmful (e.g., source memory confusion). Supernatural abilities have massive potential for VR interactions (e.g., brain-computer interfaces for telekinesis or changing laws of physics), as pointed out by Bowman et al. [16] and supported through a design method by Sadeghian and Hassenzahl [87]. Even in use cases originally meant for reproducing real experiences, such as social interactions, deviating from high realism can enrich the experience and introduce new possibilities [14, 60, 61]. Dewitz et al. [25] develop a framework on interaction techniques beyond realism, such as magic techniques, superpowers, or hyper-natural augmentation. It locates interaction techniques along the three orthogonal axes internalizability, congruence, and enhancement.

Striving for high fidelity requires time and financial effort. Therefore, it is important in research and development to critically reflect on how much realism is desirable and expedient. The IntFi Model can help identify components that should be optimized or can be less prioritized. We discuss such considerations in Subsection 8.2. While we can do a lot of good with highly realistic VR simulations, it can also be harmful and used maliciously. A growing body of literature addresses problematic implications of progressing XR technology, the ethics of increasingly attainable realism, and the risk of hostile manipulations [52, 85, 99, 108, 110], which should be considered when striving for high-fidelity applications.

Supporting the recommendation of careful tradeoffs by Jacob et al. [38] as part of their framework for reality-based interaction, we further suggest reducing realism in return for other desired qualities that align with the simulation's purpose. Jacob et al. [38] propose considering benefits in expressive power, efficiency, versatility, ergonomics, accessibility, or practicality for a tradeoff. For example, in a training simulator for learning how to juggle, high action and simulation fidelity are essential for the trainee to transfer the acquired skills to reality and apply the movements with real balls. Nevertheless, offering a training mode in slow motion to practice the movements without time pressure can be a beneficial deviation from reality.

7.3 Related Constructs

In the literature, numerous concepts and ideas have been associated with the fidelity of a simulation, such as the Place Illusion of "being there" and the Plausibility Illusion (also referred to as presence) [94, 98], coherence[93], immersion [15], engagement [50], and others [94]. We consider these as different from but correlated with fidelity. Thus, a high-fidelity interaction could result in low presence but usually leads to high presence. Conversely, a high sense of presence in a coherent, highly engaging world can also be achieved with a low-fidelity system. Slater [97] argues that high levels of place and plausibility illusions lead to realistic behavior of the user. Accordingly, the user's reactions to the virtual experience ought to correspond to how the user would react to the reference interaction, i.e., high experiential and action fidelity.

In practice, these constructs have a strong link and correlate in countless empirical studies. They also overlap in their typical assessment, e.g., some presence questionnaires comprise items to self-report perceived or experienced realism [90, 112]. Yet, the constructs concern different theoretical questions. It is, therefore, important not to confuse their claim. In particular, in empirical evaluations, the choice of measurements and interpretation of evidence depends on the concept that the research question revolves around. While there have been decades of discourse on the conception of presence and similar concepts [94], fidelity as the objective degree of correspondence between simulation and original is straightforward and with the distinct components of the IntFi Model intuitive for the planning and analysis of VR systems. Although this makes fidelity an unequivocally defined concept, it might not be the relevant one to evaluate depending on the purpose of a system or study, just as presence is not always the essential metric that should be sought after [45].

7.4 Describing Fidelity Objectively

Let's assume there is an objective, indisputable ground truth of how exact the correspondence between an original and its replication is. This truth could be described objectively if it is known. The assessments of what is true, however, can be subjective and might diverge. The more precisely we agree on how to evaluate interaction fidelity in a systematic, replicable way, the more objectively we can determine and agree on it. As discussed in the following subsection, we can only approximate the ground truth and achieve consensus through standardized measuring criteria. Technical parameters regarding system fidelity are simpler to assess objectively, while we need to rely more on subjective evaluation of user-related aspects. As various aspects determine the multi-faceted concept of interaction fidelity, it is difficult to identify its ground truth comprehensively.

Here is an example of a seemingly unambiguous and objective fidelity assessment. Probably, most people would agree that compared to the reference interaction of grasping an object, we can attribute higher interaction fidelity when the user reaches out and encloses the virtual object with their bare hand, than when the user points at the object with a hand controller and presses the trigger button. The latter implementation relies on mappings and seems less natural. However, primarily action fidelity is higher in the first implementation, while display fidelity is lower due to the lack of haptic feedback. The controller's passive force feedback has a higher correspondence to grasping a rigid object, providing higher display fidelity. Hence, depending on the focus or context, evaluations can vary. Although there is an irrefutable ground truth that we strive to ascertain, we do not necessarily succeed in doing so objectively. As a result, it is essential in communication—especially in scientific discourse—to clearly describe our perspective and reasoning.

Especially the IntFi Model's components Perceptual and Experiential Fidelity are difficult to determine objectively as the subjective nature of a person's qualia is individual and might be intrinsically subjective. Depending on the perspective on the philosophical mind–body problem, it might not even be possible to deduce mental events of the unobservable mind from the physical events in the observable brain [28]. Therefore, an objective, holistic description of interaction fidelity might be unattainable. Regardless of metaphysics and seen from a pragmatic point of view, it is more expedient to assess experiential fidelity subjectively on a user-by-user basis: Does a specific person experience the VR simulation exactly how this person would experience the reference interaction? This perspective takes all personal characteristics and biological features into account. Usually, the diversity of people's individuality must be considered because most systems are aimed at large user groups.

7.5 Measuring Fidelity

Beyond a qualitative understanding of interaction fidelity, it can be helpful to express the degree of correspondence quantitatively. Some fidelity aspects can already be assessed with high objectivity. For example, characteristics of input and output devices can be technically gauged, e.g., regarding pixel density, sensory noise, or degrees of freedom. Some of these quantifiable parameters allow a direct interpretation of how they affect interaction fidelity in direct comparisons. For example, a screen with a higher resolution than an otherwise identical screen provides higher display fidelity. For other aspects, it is harder to infer an uncontroversial effect on interaction fidelity from technical parameters. For example, regarding rendering fidelity, the influence of shaders treating light reflections differently might depend on various circumstances and is more intricate to interpret.

Some aspects are commonly assessed subjectively with self-reports, such as in questionnaires or interviews. Due to their subjective nature, experiential and perceptual fidelity are usually measured through user reports. But also some system-related components might only make sense to be assessed subjectively by large numbers of evaluators, for example, the credibility and human likeness of virtual human animations. Unfortunately, there is a limit to how much we can ask users to share their pmpressions in studies, making a holistic assessment of all subjective parameters impossible. However, we claim that it is possible to predict experiential fidelity sufficiently if enough about the other fidelity components is known.

In this work, we described the level of fidelity with the coarse categories low, medium, high, and maximum fidelity. This gives us an approximate location on the continuum and allows the rough comparison of a few systems. However, we should strive as a community for detailed, theory-based, technology-agnostic, and unambiguous metrics for all fidelity components, as outlined in the research opportunities in Section 8.

7.6 The Normative Power of Subjective Truth

There is an additional challenge when comparing VR interactions to the real world. In the case of assessing realism, we need to agree not only on the nature of the simulated interaction but also on the reality-based interaction. From a philosophical perspective, it is hard enough to agree on what "reality" objectively is. Anybody discussing the manifestation of the real world can only do so from their subjective point of view informed by their individual perceptions. While we can assess the exactness of the correspondence between a simulation and what is considered a broad consensus about the real world, the judge will ultimately be the users with their impressions. Depending on the simulation's purpose, their subjective judgment may not be decisive for how the interactions are designed, but often, experiential fidelity is the only outcome that matters and will be optimized for. In this case, only the user-related components regarding what the user perceives, experiences, and does seem important. Why should we then care for system fidelity at all? The system-related components primarily determine the levels of the user fidelity aspects. For designers and developers of VR systems and interactions, system fidelity is the only way to influence the user's perception, experience, and actions. The better we understand the components' mutual influence and interdependencies, the more effective our endeavors can be.

Interestingly, increasing the fidelity of single aspects does not necessarily increase the experiential fidelity. Previous research has suggested that reducing fidelity aspects for some interactions elicits higher experiential fidelity [59, 67]. For example, when moving through a VE using a treadmill as an input device, the walking speed is experienced as more realistic by the user when the virtual pace is exaggerated relative to the originally slower pace in the real world [10, 41, 77]. Here, the transfer function is mapped unrealistically to compensate for other limitations in interaction fidelity, such as the missing kinetic feedback from staying in place (i.e., low perceptual fidelity). The required amount of exaggeration has also been found to depend on the visual display fidelity: The smaller the field of view is, the stronger the speed must be exaggerated for the user to feel realistic [68]. Alternatively, the visual projection can be distorted to display more peripheral information [69]. Hence, reducing transfer or rendering fidelity can increase experiential fidelity. Similarly, Bowman et al. [16] argue that high fidelity in a certain component (e.g., action fidelity by rotating a Wii controller to steer a vehicle) might result in lower perceived realism because of the shortcomings in other components (in this example, missing force feedback and latency). As a consequence, the purpose of a system must be considered for prioritizing the different components' targeted level of fidelity.

7.7 Applicability to Fiction, Mixed Reality, and Other Reference Frames

The IntFi Model is designed to help assess the correspondence of interactions in virtual reality with interactions in any other reference frame. While the reference can be the real world, the model can also guide the analysis of fictional and other scenarios as long as the element to be reproduced is explicitly specified. This applies to any fictional media, imaginary narrative, dream, or fantasy, but also other VR systems, setups of previous user studies, or different levels of blended realities along the reality–virtuality continuum [96].

As an example of fictional references, the characteristics, behavior, and appearance of a lightsaber from *Star Wars* are extensively defined by the creators of the fictional artifact and, therefore, can be virtually reproduced. By contrasting the simulation with the original descriptions and depictions, we can evaluate its fidelity, as demonstrated in Example 3 of Section 4. The less clear the original to be simulated is, the more controversial the fidelity assessment might be. To stick to the Star Wars example, when simulating interactions using "the force" for telekinesis, it is less obvious how this might be realized virtually—arguably not only with a hand gesture but rather a brain-computer interface. The experiential fidelity will broadly differ between users as they have differing conceptualizations and expectations depending on what Star Wars media they have seen or read.

While developing the IntFi Model, we encountered scenarios in which no reference would be obvious to compare to, which always led us to the question: What are we trying to (re)create here? As an example from our expert interviews, participant E09 brought up an application to teach chemical processes. However, the equivalent process from reality on an atomic level made no sense to reproduce virtually for teaching. Instead, it had to be magnified to a human scale. But what could it sound like if two atoms bond on a human scale? The most suitable reference interaction we could come up with was the educator's idea of what an upscaled version of the virtualized school book model might look, sound, and feel like. This is where the model reaches its limits. A comparison using the IntFi Model provides little insight if the reference is only vaguely defined.

Another use case for applying the model is to compare two interactive systems. For example, we could compare the realism of playing baseball on a *Nintendo Wii* with an implementation in VR. While aspects of input fidelity might be similar using 6-DoF controllers, the VR version might show high fidelity in other components and explain outcomes of comparative user studies.

Similarly, the IntFi Model can be used to evaluate mixed reality (MR) applications blending virtual worlds with physical reality. There are many forms of incorporating more or less portions of different realities [96]. The interaction can be based in the real world with virtual elements integrated, or VR can be the foundation comprising elements from reality—any combination is conceivable. To apply the model meaningfully in MR contexts, it is even more important to clarify what is being compared. We advise treating the blended realities as unity and comparing it to a non-mixed equivalent. Consider, for example, an MR meeting in which co-located users participate in the real world and virtual users join remotely. To design the interactions in this blended setting, comparing them jointly to the purely real or a purely virtual equivalent can be helpful. Obviously, you can achieve high levels of fidelity when simply augmenting reality with virtual elements compared to the challenge of building a system that recreates everything virtually from scratch. Automatically, some aspects are maximum fidelity as they equal the real world. But as Lindeman and Beckhaus [49] proposed, why should we not leverage parts of the real world and augment it to create overall high-fidelity interactions if reality affords it, such as using passive haptics for teleoperation? On the other hand, we encounter limitations that are more difficult to resolve in an MR context. For example, a remote user cannot manipulate a physical object. Because the object is integrated into the interaction but not necessarily part of the system, maximum fidelity cannot be reached without virtually modifying the real world.

Another special case we would like to address is multi-user systems. Here, we find another added complexity when comparing interactions because the same encounter or activity might be experienced differently. We recommend splitting every comparison per person to avoid entangling the actions and perceptions of users or the differing hardware available to the users. Each user's interaction with the system must be considered individually for insightful analysis. This is especially important in asymmetric settings, such as in MR, where users have different possibilities and restrictions in perceiving and influencing the simulation.

7.8 Limitations

Inductively built on established HCI theory, no empirical evidence confirms the structure of the IntFi Model. The model was reviewed, practically tested, and critically discussed in interviews with 14 experts from the field, but it has not been systematically evaluated with large numbers of users in the wild. The most conclusive validation will be the community's application of the model in everyday research and development, which is yet to be seen. To provide a universal structure, our model deliberately does not include media-specific fidelity focuses such as *narrative realism* as described by Rogers et al. [84] or fidelity addressing the single senses (e.g., olfactory fidelity) but rather provides a generic framework, in which all of these can be further detailed. Depending on the use case, specific aspects can influence more than one component, such as visual fidelity, which can affect all three components of output fidelity. Another current limitation concerns the quantification of fidelity, which would help assess and compare approaches. We currently apply the approximate ranges of low, medium, high, and maximum fidelity similar to previous research [29, 50, 59] and we regard numeric assessments with standardized metrics as an opportunity for future research.

The IntFi Model is designed for examining interactions in the context of VR. This does not necessarily involve a graphical 3D environment, multimodal interfaces, a head-mounted display, or a self-representation of the user. The model can be helpful in better understanding other human-computer interactions or even non-computer-assisted technology that involves some simulation, e.g., ship navigation simulators or telemedicine interfaces. However, we emphasize that not all definitions and concepts will fit perfectly. We encourage using the IntFi Model wherever it can provide structure and guidance but advise awareness of blurred lines of systems and realities that make identifying correspondences difficult.

8 RESEARCH OPPORTUNITIES

Understanding the intricacies of building systems closely resembling the real world or other reference frames requires systematic research. The conceptual nature of the IntFi Model opens up foundational research avenues as it considers the full scope of people's interactions with virtual worlds and illustrates relations of the underlying processes. While it can be interesting to examine a component individually (e.g., action fidelity of a controller-based system), it inevitably depends on others (e.g., on display fidelity due to contact forces from holding the controller). The complexity of interdependencies, impacts on other components, and methodological challenges raise novel research questions for future work. Further, the model can inspire new perspectives on optimization and methodology.

8.1 Relationships

Which components depend on others? On the system side, there are close relations between the hardware components and the software transmitting between them, e.g., for an output device, the rendering software calculating the simulation output for a display to show must be precisely matched with both attached components. The same applies to the transfer functions translating between input devices and the computer. But also, on the user side, we find a close connection between the sensory information from the receptors and the brain processing and interpreting it. Although it is helpful for scientific discourse and system development to abstract the individual components, they cannot be regarded as independent. Even components that are not consecutive in the loop show dependencies. For instance, when assessing a system with eye tracking, the action fidelity of a user's gaze inevitably depends on rendering and display fidelity. It can only be planned or evaluated together. But how do the components generally depend on and influence each other?

Can one component compensate for another? If one aspect of realism is constrained, can an increase of another fidelity aspect make up for it? If so, can any other aspect or only a specific one? For example, in a system without a haptic display for rendering forces, modifying the control/display ratio can still induce a sense of kinesthetic forces [82, 88]. As a consequence, there is higher perceptual fidelity despite low display fidelity by deliberately lowering transfer fidelity. Are there similar compensations that allow us to build cost-effective and universal systems? Another example was given in subsection 7.6 concerning the exaggerated virtual walking speed when using a treadmill. Due to the lack of kinetic cues from not moving forward, the limitations in perceptual fidelity can be compensated by reducing transfer fidelity with a higher speed gain [67]. Similarly, if the head-mounted display's field of view is small, hence low display fidelity, a decrease in rendering fidelity due to minifying the visual output can result in higher experiential fidelity [103]. Further, when using redirected walking as a locomotion technique, action fidelity is reduced as the virtual and real paths do not match due to physical restrictions in space. We can compensate by adjusting transfer fidelity [80], simulation fidelity [105], display fidelity [9], and rendering fidelity [44] all to maintain perceptual fidelity. Nilsson et al. [67] suggest "that when limitations to a given component of fidelity reduce or distort perceptual information, then sometimes it may be possible to compensate by adjusting another component of fidelity—even if the adjustment on the surface constitutes decrease in the fidelity of the second component." Future research might investigate what other compensations should be considered in systems with restricted fidelity.

What components constitute experiential fidelity? Every component ultimately influences experiential fidelity. Some components are already well-understood through years of research, such as the aspects of input fidelity. Other components still need more scholarly attention. Above all, it requires further research to understand how strong the components' impact on experiential fidelity is. We hypothesize that the influence varies between components and cannot be reduced to a simple weighted sum of the single components. Various neural phenomena will increase complexity, such as superadditivity in multisensory integration, i.e., the effect that, for example, visual and auditory stimuli give a stronger sensory impression combined than when just adding up the individual impressions [101]. It is still unclear whether the model's components are sufficient predictors for experiential fidelity. Other influences not represented in the IntFi Model might affect perceived fidelity. For example, Witmer and Singer [112] integrate the meaningfulness of the experience in their questionnaire factor *realism*. Consequently, a dilemma would arise about how a meaningless experience from reality would be effectively simulated.

The interdisciplinary nature of this component calls for joint research, especially including psychology and cognitive science. Unraveling how experiential fidelity relates, depends, and affects related constructs, such as presence, coherence, or user experience research, is a complex endeavor that has already been embraced [94], but must be pursued in further detail—both theoretically and empirically. We suggest systematic analyses with a study design similar to the experiment by Skarbez et al. [95] on the influence of components of the plausibility illusion.

Which further patterns can be identified? Beyond the potential patterns that we proposed in Subsection 7.1, we can seek further patterns by systematically analyzing and linking empirical evidence based on the IntFi Model's structure. Similarly, most of our proposed patterns need empirical validation.

8.2 Optimization

What benefits result from improving each fidelity component? While it seems safe to assume that maximum fidelity has advantages for various user experience metrics, several studies demonstrate how less-than-perfect fidelity systems have considerable limitations in performance and preference [16]. Since it is an immensely long way to the "ultimate display" [106] indistinguishable from the real world, we currently ought to focus our research efforts on interactions with medium to high fidelity. Striving for maximum fidelity is costly and must be justified. As outlined in the Uncanny-Valley Pattern in Subsection 7.1, medium-fidelity interfaces can even result in a worse outcome than a low-fidelity implementation. Bowman et al. [16] hypothesized that hyper-natural interaction techniques (or "magic" interactions) could potentially even exceed the best possible performance of natural approaches. Yet, we need to recognize where to deviate from faithful imitation profitably.

After decades of empirical research on VR interactions, a large body of literature informs us about the effects of different levels of fidelity [12, 16, 31, 50, 78, 84, 109]. Unfortunately, it is often not further differentiated what fidelity component varies or how we can relate the findings in a broader picture. With the model's systematic and theoretical foundation to understand interaction fidelity as an overarching concept, we can better connect the dots and systematically design follow-up studies.

What are the natural limits for expedient system optimization? Technically, we could indefinitely increase interaction fidelity by reproducing reference interactions in ever greater detail, converging closer and closer to maximum fidelity. However, human perceptual sensitivity and body control are naturally limited, making further optimization futile. For example, although screen resolution could be increased to the point where we can display a grain of sand at the horizon, no user could ever tell the difference as the retina's resolution is limited. Similarly, the limited precision of performing manual tasks makes more precise tracking obsolete. Studies on such constraints in perception or body control can inform expedient system design when systematically assessed, e.g., as has been demonstrated for thresholds of redirected walking [44, 102], virtual hand offset [11, 114], latency for foveated rendering [6], or shape dissimilarity for passive haptics [24].

Should all components have a coherent level of fidelity? The discrepancy in the levels of fidelity between different components can result in low experiential fidelity, as discussed above, or bring disadvantages in performance, preference, or other user experience aspects [16, 59]. One could hypothesize that the lower fidelity is in one component of the IntFi Model, the more other components need to be enhanced to compensate. But it would also be reasonable to assume that other components must be matched at the same level to give the user a consistent impression. For example, in Mario Kart with cartoony visuals and a comical setting, high-fidelity vehicle control and physics would seem inappropriate, might decrease the perceived overall fidelity, and limit performance. Instead, the transfer functions and car behavior that make driving simple and error-tolerant lead to a coherent experience and arguably higher experiential fidelity. Similarly, we hypothesize that a driving simulator with sophisticated car physics and true-to-life input devices will be experienced as most realistic if the sensory feedback matches the high faithfulness and does not rely on cartoony visuals, funny sounds, or lacks haptics. Systematic experimental comparisons might reveal how uniform interactions should be designed.

8.3 Methodology

How can we quantify fidelity? Ideally, every component of the IntFi Model would come with means of quantification or a theoretically founded metric. Due to the scope of this universal model and the depth of its conceivable subcomponents, achieving a set of methods for quantifying all fidelity components comprehensively is a considerable endeavor that the research community has worked on and will arguably need to continue working on for decades. Al-Jundi and Tanbour [5] proposed five categories in their framework for evaluating some fidelity aspects. The limitations of human sensory capabilities partially define the maximum. The other classifications are not delimited clearly. It is challenging to evaluate the moving target of rapidly evolving technology as it requires either dynamic adjustment of the classification or prospective universality. Ideally, a comprehensive framework would allow assigning numeric values objectively, reproducibly, and universally. Looking at haptic fidelity as an example, the Haptic Fidelity Framework by Muender et al. [64] shows how complex and manifold it can be to specify even one of the subcomponents of display fidelity. This specialized framework identifies 14 factors defining haptic fidelity along the categories *sensing*, *hardware*, and *software*. The publication includes an expert tool to quantify each factor and calculate an aggregated haptic fidelity score on a five-point Likert scale for technology-agnostic comparison. Considering that this covers only one modality in one out of eight fidelity components, it poses a significant research opportunity to provide such sophisticated frameworks for all aspects of interaction fidelity. Meanwhile, a validated questionnaire for users' self-reports in studies on haptic fidelity is still missing—leading to the following research opportunity.

How can we measure each aspect? Given we have means to describe all fidelity factors quantitatively, we need to establish methods, standards, and instruments to measure it objectively and subjectively. Gonçalves et al. [30] recently presented a systematic literature review on the methodology of 79 studies on VR realism. For heuristic expert analysis, specialized frameworks, such as the FIFA [59], the Haptic Fidelity Framework[64], or the Simulation Fidelity Rating Scale for flight simulators [74] can help investigate single (sub)components, as outlined before. Furthermore, standards for technical evaluations are needed to compare devices, for instance, regarding the physical similarity of generated sensory stimuli or the accuracy of tracking devices. Validated instruments for psychometric evaluations allow comparison between studies on perceptual fidelity. Case-specific behavioral measures can improve our assessment of how realistically users react but are difficult to standardize. For subjective assessments, specialized and validated possibilities for self-reporting are essential beyond broad subscales of presence questionnaires. We propose developing dedicated tools to understand the perceived fidelity of the distinct interface components instead of generalizing overall realism.

Which specialized tools are out there? Gathering the already available resources on fidelity research would significantly support the community. However, comprehensively collecting and arranging all frameworks, instruments, questionnaires, etc., is challenging. This work provides a first step, particularly with the overview in Table 2. Beyond that, a systematic review is needed.

What should reporting guidelines include? Another crucial challenge for interaction fidelity research is finding standard reporting guidelines for fostering comparability and generalizability. Currently, relevant information about the system design is often missing to understand evaluation results and apply meta-analyses comprehensively. The IntFi Model might serve as a starting point to agree on reporting guidelines ensuring all system components that define interaction fidelity as a holistic concept are being reported. Because of the interdependencies of the single aspects, we encourage researchers to report details about the interactions beyond the element of interest.

9 CONCLUSION

To understand what makes interactions in VR simulations more or less faithful to the real world or any other reference frame, we must distinguish between various aspects of fidelity. In this article, we proposed the Interaction Fidelity Model (IntFi Model) that allows analyzing how closely a virtual interaction corresponds to the original along various factors. We define eight fidelity components along the HCI loop: action, detection, transfer, simulation, rendering, display, perceptual, and experiential fidelity.

The consequent terminology offered in this work supports precise communication and consistency across publications in the field. With a clear structure, rigorous explanations, practical examples, and a guideline with best practices, we demonstrate how VR professionals in research and development can use the conceptual model to describe, understand, compare, hypothesize, and teach. With its theoretically grounded simplicity, the IntFi Model can be universally applied to any VR experience. Therefore, our taxonomy defines only twelve general fidelity terms, as listed in Table 1. Beyond that, this article serves as a signpost referring readers to previous publications with specialized frameworks or concepts, as each component can be further distinguished in more detail.

The presented model underwent rigorous, critical discussions and was refined iteratively. For validation, we conducted 14 extensive interviews with experts from academia and the industry to review and test the model from different perspectives. The thematic analysis showed criticism of the concept, the terminology, and applications, identified application strategies, and outlined various benefits and use cases of the model. All experts found it interesting and helpful. As suggested in the interviews, we provide educational material as part of the supplemental material, including modern posters and a slide deck, which are free to use and adapt.

From our practical experiences with the model, we identified common patterns that might be prevalent in various use cases and interaction techniques. By connecting the dots in such a way, the IntFi Model will support finding similarities in study results and see their findings in a bigger picture. Using the model to think about the fidelity of VR interactions opens up promising opportunities for systematic and targeted research. We hope to inspire new directions in research for a better understanding of interactions in VR.

ACKNOWLEDGMENTS

We are grateful to the experts for sharing their perspectives and criticism in the interviews. We thank Yvonne Rogers, Don Norman, and Katja Rogers for helping shape this work, and Laura Canil for realizing the visual designs of Figures 1 and 3.

Funding

This research was partially funded by the Klaus Tschira Stiftung.

Conflicts of Interest

The authors report there are no competing interests to declare.

REFERENCES

[1] Gregory D Abowd and Russell Beale. 1991. Users, systems and interfaces: A unifying framework for interaction. In *HCI 1991: People and Computers IV*. Vol. 4. 73–87.

[2] Parastoo Abtahi, Sidney Q Hough, James A Landay, and Sean Follmer. 2022. Beyond Being Real: A Sensorimotor Control Perspective on Interactions in Virtual Reality. In *CHI Conference on Human Factors in Computing Systems*. 1–17.

[3] Ernest Adams. 2014. *Fundamentals of Game Design*. New Riders.

[4] Jindřich Adolf, Peter Kán, Benjamin Outram, Hannes Kaufmann, Jaromír Doležal, and Lenka Lhotská. 2019. Juggling in VR: Advantages of Immersive Virtual Reality in Juggling Learning. In *Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology* (New York, NY, USA, 2019-11-12) *(VRST '19)*. Association for Computing Machinery, 1–5. https://doi.org/10.1145/3359996.3364246

[5] Hamza A Al-Jundi and Emad Y Tanbour. 2022. A framework for fidelity evaluation of immersive virtual reality systems. *Virtual Reality* (2022), 1–20.

[6] Rachel Albert, Anjul Patney, David Luebke, and Joohwan Kim. 2017. Latency Requirements for Foveated Rendering in Virtual Reality. *ACM Transactions on Applied Perception* 14, 4 (Sept. 2017), 25:1–25:13. https://doi.org/10.1145/3127589

[7] Amy L Alexander, Tad Brunyé, Jason Sidman, Shawn A Weil, et al. 2005. From gaming to training: A review of studies on fidelity, immersion, presence, and buy-in and their effects on transfer in pc-based simulations and games. *DARWARS Training Impact Group* 5 (2005), 1–14.

[8] Tom Arthur, Tilly Loveland-Perkins, Charlotte Williams, David Harris, Mark Wilson, Toby de Burgh, Jagtar Dhanda, and Sam Vine. 2023. Examining the validity and fidelity of a virtual reality simulator for basic life support training. *BMC Digital Health* 1, 1 (2023), 16. https://doi.org/10.1186/s44247-023-00016-1

[9] Jonas Auda, Max Pascher, and Stefan Schneegass. 2019. Around the (Virtual) World: Infinite Walking in Virtual Reality Using Electrical Muscle Stimulation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3290605.3300661

[10] Tom Banton, Jeanine Stefanucci, Frank Durgin, Adam Fass, and Dennis Proffitt. 2005. The perception of walking speed in a virtual environment. *Presence: Teleoperators and Virtual Environments* 14, 4 (Aug. 2005), 394–406. https://doi.org/10.1162/105474605774785262

[11] Brett Benda, Shaghayegh Esmaeili, and Eric D. Ragan. 2020. Determining Detection Thresholds for Fixed Positional Offsets for Virtual Hand Remapping in Virtual Reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE Computer Society, 269–278. https://doi.org/10.1109/ISMAR50242.2020.00050

[12] Ayush Bhargava, Jeffrey W. Bertrand, Anand K. Gramopadhye, Kapil C. Madathil, and Sabarish V. Babu. 2018. Evaluating Multiple Levels of an Interaction Fidelity Continuum on Performance and Learning in Near-Field Training Simulations. *IEEE Transactions on Visualization and Computer Graphics* 24, 4 (April 2018), 1418–1427. https://doi.org/10.1109/TVCG.2018.2794639

[13] Michael Bonfert, Stella Lemke, Robert Porzel, and Rainer Malaka. 2022. Kicking in Virtual Reality: The Influence of Foot Visibility on the Shooting Experience and Accuracy. In *Proceedings of the 2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (Virtual, 2022-03-13). IEEE Computer Society, 711–718. https://doi.org/10.1109/VR51125.2022.00092

[14] Michael Bonfert, Anke V. Reinschluessel, Susanne Putze, Yenchin Lai, Dmitry Alexandrovsky, Rainer Malaka, and Tanja Döring. 2023. Seeing the faces is so important—Experiences from online team meetings on commercial virtual reality platforms. *Frontiers in Virtual Reality* 3 (2023), 19. https://doi.org/10.3389/frvir.2022.945791

[15] Doug A. Bowman and Ryan P. McMahan. 2007. Virtual Reality: How Much Immersion Is Enough? *Computer* 40, 7 (July 2007), 36–43. https://doi.org/10.1109/MC.2007.257

[16] Doug A. Bowman, Ryan P. McMahan, and Eric D. Ragan. 2012. Questioning naturalism in 3D user interfaces. *Commun. ACM* 55, 9 (Sept. 2012), 78–88. https://doi.org/10.1145/2330667.2330687

[17] Gerd Bruder, Paul Lubos, and Frank Steinicke. 2015. Cognitive Resource Demands of Redirected Walking. *IEEE Transactions on Visualization and Computer Graphics* 21, 4 (April 2015), 539–544. https://doi.org/10.1109/TVCG.2015.2391864

[18] Dan Calatayud, Sonal Arora, Rajesh Aggarwal, Irina Kruglikova, Svend Schulze, Peter Funch-Jensen, and Teodor Grantcharov. 2010. Warm-up in a virtual reality environment improves performance in the operating room. *Annals of surgery* 251, 6 (2010), 1181–1185.

[19] Stephanie Carnell, Anna Miles, and Benjamin Lok. 2022. Evaluating Virtual Patient Interaction Fidelity With Advanced Communication Skills Learners. *Frontiers in Virtual Reality* 2 (2022). https://doi.org/10.3389/frvir.2021.801793

[20] Vuthea Chheang, Patrick Saalfeld, Tobias Huber, Florentine Huettl, Werner Kneist, Bernhard Preim, and Christian Hansen. 2019. Collaborative virtual reality for laparoscopic liver surgery training. In *2019 IEEE International Conference*

on Artificial Intelligence and Virtual Reality (AIVR). IEEE, 1–17.
[21] Peter H Cosman, Thomas J Hugh, Christopher J Shearer, Neil D Merrett, Andrew V Biankin, and John A Cartmill. 2007. Skills acquired on virtual reality laparoscopic simulators transfer into the operating room in a blinded, randomised, controlled trial.. In *MMVR*. 76–81.
[22] C. Cox, B. Hicks, and J. Gopsill. 2022. Improving Mixed-Reality Prototyping through a Classification and Characterisation of Fidelity. *Proceedings of the Design Society* 2 (May 2022), 353–362. https://doi.org/10.1017/pds.2022.37
[23] Xavier de Tinguy, Thomas Howard, Claudio Pacchierotti, Maud Marchal, and Anatole Lécuyer. 2020. WeATaViX: WEarable Actuated TAngibles for VIrtual Reality eXperiences. In *Haptics: Science, Technology, Applications* (Cham, 2020) *(Lecture Notes in Computer Science)*, Ilana Nisky, Jess Hartcher-O'Brien, Michaël Wiertlewski, and Jeroen Smeets (Eds.). Springer International Publishing, 262–270. https://doi.org/10.1007/978-3-030-58147-3_29
[24] Xavier de Tinguy, Claudio Pacchierotti, Mathieu Emily, Mathilde Chevalier, Aurelie Guignardat, Morgan Guillaudeux, Chloe Six, Anatole Lecuyer, and Maud Marchal. 2019. How Different Tangible and Virtual Objects Can Be While Still Feeling the Same?. In *2019 IEEE World Haptics Conference (WHC)*. IEEE, Tokyo, Japan, 580–585. https://doi.org/10.1109/WHC.2019.8816164
[25] Bastian Dewitz, Sukran Karaosmanoglu, Robert W. Lindeman, and Frank Steinicke. 2023. 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)*. 807–808. https://doi.org/10.1109/VRW58643.2023.00246
[26] Andrew Feinstein and Hugh Cannon. 2001. Fidelity, Verifiability, and Validity of Simulation: Constructs for Evaluation. *Developments in Business Simulation and Experiential Learning* 28 (April 2001), 57–67.
[27] Daniel Freeman, Polly Haselton, Jason Freeman, Bernhard Spanlang, Sameer Kishore, Emily Albery, Megan Denne, Poppy Brown, Mel Slater, and Alecia Nickless. 2018. Automated psychological therapy using immersive virtual reality for treatment of fear of heights: a single-blind, parallel-group, randomised controlled trial. *The Lancet Psychiatry* 5, 8 (2018), 625–632.
[28] Danko D. Georgiev. 2020. Quantum information theoretic approach to the mind–brain problem. *Progress in Biophysics and Molecular Biology* 158 (Dec. 2020), 16–32. https://doi.org/10.1016/j.pbiomolbio.2020.08.002
[29] Stephen B. Gilbert. 2016. Perceived Realism of Virtual Environments Depends on Authenticity. *Presence: Teleoperators and Virtual Environments* 25, 4 (Dec. 2016), 322–324. https://doi.org/10.1162/PRES_a_00276
[30] Guilherme Gonçalves, Pedro Monteiro, Hugo Coelho, Miguel Melo, and Maximino Bessa. 2021. Systematic Review on Realism Research Methodologies on Immersive Virtual, Augmented and Mixed Realities. *IEEE Access* 9 (2021), 89150–89161. https://doi.org/10.1109/ACCESS.2021.3089946
[31] Tavinor Grant. 2019. Towards an analysis of virtual realism. In *DiGRA &#3919 - Proceedings of the 2019 DiGRA International Conference: Game, Play and the Emerging Ludo-Mix*. DiGRA. http://www.digra.org/wp-content/uploads/digital-library/DiGRA_2019_paper_24.pdf
[32] David J. Harris, Jonathan M. Bird, Philip A. Smart, Mark R. Wilson, and Samuel J. Vine. 2020. A Framework for the Testing and Validation of Simulated Environments in Experimentation and Training. *Frontiers in Psychology* 11 (2020). https://www.frontiersin.org/articles/10.3389/fpsyg.2020.00605
[33] Casper Harteveld. 2011. *Triadic Game Design: Balancing Reality, Meaning and Play*. Springer, London. https://doi.org/10.1007/978-1-84996-157-8
[34] R. Hays and M. Singer. 1989. *Simulation Fidelity in Training System Design: Bridging the Gap Between Reality and Training*. Springer, New York, NY, USA.
[35] Robert T. Hays, John W. Jacobs, Carolyn Prince, and Eduardo Salas. 1992. Flight Simulator Training Effectiveness: A Meta-Analysis. *Military Psychology* 4, 2 (June 1992), 63–74. https://doi.org/10.1207/s15327876mp0402_1
[36] Shirley Holt. 2023. Virtual reality, augmented reality and mixed reality: For astronaut mental health; and space tourism, education and outreach. *Acta Astronautica* 203 (Feb. 2023), 436–446. https://doi.org/10.1016/j.actaastro.2022.12.016
[37] Kasper Hornbæk and Antti Oulasvirta. 2017. What Is Interaction?. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 5040–5052. https://doi.org/10.1145/3025453.3025765
[38] Robert JK Jacob, Audrey Girouard, Leanne M Hirshfield, Michael S Horn, Orit Shaer, Erin Treacy Solovey, and Jamie Zigelbaum. 2008. Reality-based interaction: a framework for post-WIMP interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 201–210.
[39] Joseph J. Jr. LaViola, Ernst Kruijff, Ryan P. McMahan, Doug A. Bowman, and Ivan Poupyrev. 2017. *3D User Interfaces: Theory and Practice* (2 ed.). Addison Wesley, Boston.
[40] Bill Kapralos, Fuad Moussa, and Adam Dubrowski. 2014. An Overview of Virtual Simulation and Serious Gaming for Surgical Education and Training. In *Technologies of Inclusive Well-Being: Serious Games, Alternative Realities, and Play Therapy*, Anthony Lewis Brooks, Sheryl Brahnam, and Lakhmi C. Jain (Eds.). Springer, Berlin, Heidelberg, 289–306. https://doi.org/10.1007/978-3-642-45432-5_14
[41] Laura Kassler, Jeff Feasel, Michael D. Lewek, Frederick P. Brooks, and Mary C. Whitton. 2010. Matching actual treadmill walking speed and visually perceived walking speed in a projection virtual environment. In *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization (APGV '10)*. Association for Computing Machinery, New York, NY, USA, 161. https://doi.org/10.1145/1836248.1836283
[42] Erin Kim and Oliver Schneider. 2020. Defining haptic experience: Foundations for understanding, communicating, and evaluating HX. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.
[43] Robert Kovacs, Eyal Ofek, Mar Gonzalez Franco, Alexa Fay Siu, Sebastian Marwecki, Christian Holz, and Mike Sinclair. 2020. Haptic PIVOT: On-Demand Handhelds in VR. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (Virtual Event USA, 2020-10-20). ACM, 1046–1059. https://doi.org/10.1145/3379337.3415854
[44] Eike Langbehn, Frank Steinicke, Markus Lappe, Gregory F. Welch, and Gerd Bruder. 2018. In the blink of an eye: leveraging blink-induced suppression for imperceptible position and orientation redirection in virtual reality. *ACM Transactions on Graphics* 37, 4 (Aug. 2018), 1–11. https://doi.org/10.1145/3197517.3201335
[45] Marc Erich Latoschik and Carolin Wienrich. [n. d.]. Congruence and Plausibility, Not Presence: Pivotal Conditions for XR Experiences and Effects, a Novel Approach. 3 ([n. d.]). https://doi.org/10.3389/frvir.2022.694433
[46] Susan J Lederman and Roberta L Klatzky. 2009. Haptic perception: A tutorial. *Attention, Perception, & Psychophysics* 71, 7 (2009), 1439–1459.
[47] Alexander Lindau, Vera Erbes, Steffen Lepa, Hans-Joachim Maempel, Fabian Brinkman, and Stefan Weinzierl. 2014. A spatial audio quality inventory (SAQI). *Acta Acustica united with Acustica* 100, 5 (2014), 984–994.
[48] Robert W. Lindeman and Steffi Beckhaus. 2009. Crafting Memorable VR Experiences Using Experiential Fidelity. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology* (Kyoto, Japan) *(VRST '09)*. Association for Computing Machinery, New York, NY, USA, 187–190. https://doi.org/10.1145/1643928.1643970
[49] Robert W. Lindeman and Steffi Beckhaus. 2009. Crafting memorable VR experiences using experiential fidelity. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology* (New York, NY, USA, 2009-11-18) *(VRST '09)*. Association for Computing Machinery, 187–190. https://doi.org/10.1145/1643928.1643970
[50] Heide Lukosch, Stephan Lukosch, Simon Hoermann, and Robert W. Lindeman. 2019. Conceptualizing Fidelity for HCI in Applied Gaming. In *HCI in Games - 1st International Conference, HCI-Games 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Proceedings (LNCS)*, Xiaowen Fang (Ed.). Springer Nature Switzerland AG, 165–179. https://doi.org/10.1007/978-3-030-22602-2_14
[51] Ke Ma and Bernhard Hommel. 2015. The role of agency for perceived ownership in the virtual hand illusion. *Consciousness and Cognition* 36 (2015), 277–288. https://doi.org/10.1016/j.concog.2015.07.008
[52] Michael Madary and Thomas K. Metzinger. 2016. Real Virtuality: A Code of Ethical Conduct. Recommendations for Good Scientific Practice and the Consumers of VR-Technology. *Frontiers in Robotics and AI* 3 (2016). https://www.frontiersin.org/articles/10.3389/frobt.2016.00003
[53] K. Mania, D. Wooldridge, M. Coxon, and A. Robinson. 2006. The effect of visual and interaction fidelity on spatial cognition in immersive virtual environments. *IEEE Transactions on Visualization and Computer Graphics* 12, 3 (2006), 396–404. https://doi.org/10.1109/TVCG.2006.55
[54] Mori Masahiro. 1970. The uncanny valley. *Energy* 7, 4 (1970), 33–35.
[55] Ryan Patrick McMahan. 2011. *Exploring the effects of higher-fidelity display and interaction for virtual reality games*. Ph. D. Dissertation. Virginia Tech.
[56] Ryan P. McMahan. 2018. Virtual Reality System Fidelity. In *Encyclopedia of Computer Graphics and Games*, Newton Lee (Ed.). Springer International Publishing, Cham, 1–8. https://doi.org/10.1007/978-3-319-08234-9_251-1
[57] Ryan P. McMahan, Alexander Joel D. Alon, Shaimaa Lazem, Robert J. Beaton, David Machaj, Michael Schaefer, Mara G. Silva, Anamary Leal, Robert Hagan, and Doug A. Bowman. 2010. Evaluating natural interaction techniques in video games. In *2010 IEEE Symposium on 3D User Interfaces (3DUI)*. 11–14. https://doi.org/10.1109/3DUI.2010.5444727
[58] Ryan P. McMahan, Doug A. Bowman, David J. Zielinski, and Rachael B. Brady. 2012. Evaluating Display Fidelity and Interaction Fidelity in a Virtual Reality Game. *IEEE Transactions on Visualization and Computer Graphics* 18, 4 (April 2012), 626–633. https://doi.org/10.1109/TVCG.2012.43
[59] Ryan P. McMahan, Chengyuan Lai, and Swaroop K. Pal. 2016. Interaction Fidelity: The Uncanny Valley of Virtual Reality Interactions. In *Virtual, Augmented and Mixed Reality (Lecture Notes in Computer Science)*, Stephanie Lackey and Randall Shumaker (Eds.). Springer International Publishing, Cham, 59–70. https://doi.org/10.1007/978-3-319-39907-2_6
[60] Joshua McVeigh-Schultz and Katherine Isbister. 2021. The Case for "Weird Social" in VR/XR: A Vision of Social Superpowers Beyond Meatspace. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, 1–10.

https://doi.org/10.1145/3411763.3450377
[61] Joshua McVeigh-Schultz and Katherine Isbister. 2022. A "beyond being there" for VR meetings: envisioning the future of remote work. *Human–Computer Interaction* 37, 5 (Sept. 2022), 433–453. https://doi.org/10.1080/07370024.2021.1994860
[62] Daniel Medeiros, Maurıcio Sousa, Alberto Raposo, and Joaquim Jorge. 2020. Magic Carpet: Interaction Fidelity for Flying in VR. *IEEE Transactions on Visualization and Computer Graphics* 26, 9 (2020), 2793–2804. https://doi.org/10.1109/TVCG.2019.2905200
[63] Merriam-Webster. 2024. Definition of Fidelity. https://www.merriam-webster.com/dictionary/fidelity
[64] Thomas Muender, Michael Bonfert, Anke Verena Reinschluessel, Rainer Malaka, and Tanja Döring. 2022. Haptic Fidelity Framework: Defining the Factors of Realistic Haptic Feedback for Virtual Reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 431, 17 pages. https://doi.org/10.1145/3491102.3501953
[65] Mahdi Nabiyouni. 2017. *How Does Interaction Fidelity Influence User Experience in VR Locomotion?* Doctoral Dissertation. https://vtechworks.lib.vt.edu/handle/10919/74945
[66] Mahdi Nabiyouni, Ayshwarya Saktheeswaran, Doug A. Bowman, and Ambika Karanth. 2015. Comparing the performance of natural, semi-natural, and non-natural locomotion techniques in virtual reality. In *2015 IEEE Symposium on 3D User Interfaces (3DUI)*. 3–10. https://doi.org/10.1109/3DUI.2015.7131717
[67] Niels Nilsson, Rolf Nordahl, and Stefania Serafin. 2017. *Waiting for the Ultimate Display: Can Decreased Fidelity Positively Influence Perceived Realism?* https://doi.org/10.1109/WEVR.2017.7957710
[68] Niels Christian Nilsson, Stefania Serafin, and Rolf Nordahl. 2014. Establishing the Range of Perceptually Natural Visual Walking Speeds for Virtual Walking-In-Place Locomotion. *IEEE Transactions on Visualization and Computer Graphics* 20, 4 (April 2014), 569–578. https://doi.org/10.1109/TVCG.2014.21
[69] Niels Christian Nilsson, Stefania Serafin, and Rolf Nordahl. 2015. The effect of visual display properties and gain presentation mode on the perceived naturalness of virtual walking speeds. In *2015 IEEE Virtual Reality (VR)*. 81–88. https://doi.org/10.1109/VR.2015.7223328
[70] Donald A. Norman. 1989. *The Design of Everyday Things*. Doubleday.
[71] Donald A. Norman and Stephen W. Draper. 1986. *User centered system design: new perspectives on human-computer interaction*. L. Erlbaum, Hillsdale (N. J.) London.
[72] Philippe Palanque, Andy Cockburn, and Carl Gutwin. 2020. A Classification of Faults Covering the Human-Computer Interaction Loop. In *Computer Safety, Reliability, and Security (Lecture Notes in Computer Science)*, António Casimiro, Frank Ortmeier, Friedemann Bitsch, and Pedro Ferreira (Eds.). Springer International Publishing, Cham, 434–448. https://doi.org/10.1007/978-3-030-54549-9_29
[73] Randy Pausch, Jon Snoddy, Robert Taylor, Scott Watson, and Eric Haseltine. 1996. Disney's Aladdin: first steps toward storytelling in virtual reality. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques (SIGGRAPH '96)*. Association for Computing Machinery, New York, NY, USA, 193–203. https://doi.org/10.1145/237170.237257
[74] P. Perfect, E. Timson, M. D. White, G. D. Padfield, R. Erdos, and A. W. Gubbels. 2014. A rating scale for the subjective assessment of simulation fidelity. *The Aeronautical Journal* 118, 1206 (2014), 953–974. https://doi.org/10.1017/S0001924000009635
[75] Sandra Poeschl. 2017. Virtual reality training for public speaking—a QUEST-VR framework validation. *Frontiers in ICT* 4 (2017), 13.
[76] Ivan Poupyrev, Mark Billinghurst, Suzanne Weghorst, and Tadao Ichikawa. 1996. 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* (New York, NY, USA, 1996-11-01) *(UIST '96)*. Association for Computing Machinery, 79–80. https://doi.org/10.1145/237091.237102
[77] Wendy Powell, Brett Stevens, Steve Hand, and M. Simmonds. 2011. Blurring the boundaries: the perception of visual gain in treadmill-mediated virtual environments: 3rd IEEE VR 2011 Workshop on Perceptual Illusions in Virtual Environments. In *3rd IEEE VR 2011 Workshop on Perceptual Illusions in Virtual Environments*. IEEE, Singapore.
[78] Eric D. Ragan, Doug A. Bowman, Regis Kopper, Cheryl Stinson, Siroberto Scerbo, and Ryan P. McMahan. 2015. Effects of Field of View and Visual Complexity on Virtual Reality Training Effectiveness for a Visual Scanning Task. *IEEE Transactions on Visualization and Computer Graphics* 21, 7 (July 2015), 794–807. https://doi.org/10.1109/TVCG.2015.2403312
[79] John R. Raser. 1969. *Simulation and Society: An Exploration of Scientific Gaming*. Allyn and Bacon.
[80] Sharif Razzaque, Zachariah Kohn, and Mary C. Whitton. 2001. Redirected Walking. In *Eurographics 2001 - Short Presentations*. Eurographics Association. https://doi.org/10.2312/egs.20011036
[81] Trygve Reenskaug. 1979. *The original MVC reports*. Technical Report. Xerox Palo Alto Research Laboratory (PARC), Palo Alto, CA. 14 pages. https://folk.universitetetioslo.no/trygver/2007/MVC_Originals.pdf
[82] Michael Rietzler, Florian Geiselhart, Julian Frommel, and Enrico Rukzio. 2018. Conveying the Perception of Kinesthetic Feedback in Virtual Reality using State-of-the-Art Hardware. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. ACM Press, Montreal QC, Canada, 1–13. https://doi.org/10.1145/3173574.3174034
[83] Katja Rogers, Jana Funke, Julian Frommel, Sven Stamm, and Michael Weber. 2019. Exploring Interaction Fidelity in Virtual Reality: Object Manipulation and Whole-Body Movements. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300644
[84] Katja Rogers, Sukran Karaosmanoglu, Maximilian Altmeyer, Ally Suarez, and Lennart E. Nacke. 2022. Much Realistic, Such Wow! A Systematic Literature Review of Realism in Digital Games. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 190, 21 pages. https://doi.org/10.1145/3491102.3501875
[85] Marius Rubo, Nadine Messerli, and Simone Munsch. 2021. The human source memory system struggles to distinguish virtual reality and reality. *Computers in Human Behavior Reports* 4 (Aug. 2021), 100111. https://doi.org/10.1016/j.chbr.2021.100111
[86] Stuart Russell and Peter Norvig. 2020. *Artificial Intelligence: A Modern Approach (4th Edition)*. Pearson. http://aima.cs.berkeley.edu/
[87] Shadan Sadeghian and Marc Hassenzahl. 2021. 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* (Virtual Event, USA) *(DIS '21)*. Association for Computing Machinery, New York, NY, USA, 180–189. https://doi.org/10.1145/3461778.3462111
[88] Majed Samad, Elia Gatti, Anne Hermes, Hrvoje Benko, and Cesare Parise. 2019. Pseudo-Haptic Weight: Changing the Perceived Weight of Virtual Objects By Manipulating Control-Display Ratio. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300550
[89] Daniel Sarkady, Larissa Neuburger, and Roman Egger. 2021. Virtual reality as a travel substitution tool during COVID-19. In *Information and communication technologies in tourism 2021*. Springer, 452–463.
[90] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. 2001. The Experience of Presence: Factor Analytic Insights. *Presence: Teleoperators and Virtual Environments* 10, 3 (June 2001), 266–281. https://doi.org/10.1162/105474601300343603
[91] Peter Schulz, Dmitry Alexandrovsky, Felix Putze, Rainer Malaka, and Johannes Schöning. 2019. The role of physical props in vr climbing environments. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.
[92] Stefania Serafin, Ali Adjorlu, Niels Nilsson, Lui Thomsen, and Rolf Nordahl. 2017. Considerations on the use of virtual and augmented reality technologies in music education. In *2017 IEEE virtual reality workshop on K-12 embodied learning through virtual & augmented reality (KELVAR)*. IEEE, 1–4.
[93] Richard Skarbez. 2016. *Plausibility Illusion in Virtual Environments*. Doctoral Dissertation. University of North Carolina at Chapel Hill Graduate School, Chapel Hill. https://doi.org/10.17615/2mg3-gh93
[94] Richard Skarbez, Frederick P. Brooks, Jr., and Mary C. Whitton. 2017. A Survey of Presence and Related Concepts. *Comput. Surveys* 50, 6 (Nov. 2017), 96:1–96:39. https://doi.org/10.1145/3134301
[95] Richard Skarbez, Solène Neyret, Frederick P. Brooks, Mel Slater, and Mary C. Whitton. 2017. A Psychophysical Experiment Regarding Components of the Plausibility Illusion. *IEEE Transactions on Visualization and Computer Graphics* 23, 4 (2017), 1369–1378. https://doi.org/10.1109/TVCG.2017.2657158
[96] Richard Skarbez, Missie Smith, and Mary C. Whitton. 2021. Revisiting Milgram and Kishino's Reality-Virtuality Continuum. *Frontiers in Virtual Reality* 2 (2021). https://doi.org/10.3389/frvir.2021.647997
[97] Mel Slater. 2009. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364, 1535 (Dec. 2009), 3549–3557. https://doi.org/10.1098/rstb.2009.0138
[98] Mel Slater, Domna Banakou, Alejandro Beacco, Jaime Gallego, Francisco Macia-Varela, and Ramon Oliva. 2022. A Separate Reality: An Update on Place Illusion and Plausibility in Virtual Reality. *Frontiers in Virtual Reality* 3 (2022). https://www.frontiersin.org/articles/10.3389/frvir.2022.914392
[99] Mel Slater, Cristina Gonzalez-Liencres, Patrick Haggard, Charlotte Vinkers, Rebecca Gregory-Clarke, Steve Jelley, Zillah Watson, Graham Breen, Raz Schwarz, William Steptoe, Dalila Szostak, Shivashankar Halan, Deborah Fox, and Jeremy Silver. 2020. The Ethics of Realism in Virtual and Augmented Reality. *Frontiers in Virtual Reality* 1 (2020). https://www.frontiersin.org/articles/10.3389/frvir.2020.00001
[100] Mel Slater and Sylvia Wilbur. 1997. A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments.

Presence: Teleoperators & Virtual Environments 6, 6 (1997), 603–616.

[101] Terrence R. Stanford and Barry E. Stein. 2007. Superadditivity in multisensory integration: putting the computation in context. *NeuroReport* 18, 8 (May 2007), 787–792. https://doi.org/10.1097/WNR.0b013e3280c1e315

[102] Frank Steinicke, Gerd Bruder, Jason Jerald, Harald Frenz, and Markus Lappe. 2010. Estimation of Detection Thresholds for Redirected Walking Techniques. *IEEE Transactions on Visualization and Computer Graphics* 16, 1 (Jan. 2010), 17–27. https://doi.org/10.1109/TVCG.2009.62

[103] Frank Steinicke, Gerd Bruder, Scott Kuhl, Pete Willemsen, Markus Lappe, and Klaus Hinrichs. 2011. Natural Perspective Projections for Head-Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics* 17, 7 (July 2011), 888–899. https://doi.org/10.1109/TVCG.2010.248

[104] Thomas A Stoffregen, Benoit G Bardy, LJ Smart, and RJ Pagulayan. 2003. On the nature and evaluation of fidelity in virtual environments. *Virtual and adaptive environments: Applications, implications, and human performance issues* (2003), 111–128.

[105] Evan A. Suma, Zachary Lipps, Samantha Finkelstein, David M. Krum, and Mark Bolas. 2012. Impossible Spaces: Maximizing Natural Walking in Virtual Environments with Self-Overlapping Architecture. *IEEE Transactions on Visualization and Computer Graphics* 18, 4 (April 2012), 555–564. https://doi.org/10.1109/TVCG.2012.47

[106] Ivan E. Sutherland. 1965. The Ultimate Display. In *Proceedings of the IFIP Congress*. 506–508.

[107] Kaisa Tiippana. 2014. What is the McGurk effect? *Frontiers in Psychology* 5 (2014). https://www.frontiersin.org/articles/10.3389/fpsyg.2014.00725

[108] Wen-Jie Tseng, Elise Bonnail, Mark McGill, Mohamed Khamis, Eric Lecolinet, Samuel Huron, and Jan Gugenheimer. 2022. The Dark Side of Perceptual Manipulations in Virtual Reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3491102.3517728

[109] Marnix van Gisbergen, Michelle Kovacs, Fabio Campos, Malou van der Heeft, and Valerie Vugts. 2019. What We Don't Know. The Effect of Realism in Virtual Reality on Experience and Behaviour. In *Augmented Reality and Virtual Reality*, M. Claudia tom Dieck and Timothy Jung (Eds.). Springer International Publishing, Cham, 45–57. https://doi.org/10.1007/978-3-030-06246-0_4

[110] Brian D. Wassom. 2015. *Augmented reality law, privacy, and ethics: law, society, and emerging AR technologies*. Syngress, Waltham, MA.

[111] J. Williamson, R. Murray-Smith, B. Blankertz, M. Krauledat, and K. R. Müller. 2009. Designing for uncertain, asymmetric control: Interaction design for brain–computer interfaces. *International Journal of Human-Computer Studies* 67, 10 (Oct. 2009), 827–841. https://doi.org/10.1016/j.ijhcs.2009.05.009

[112] Bob G. Witmer and Michael J. Singer. 1998. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoper. Virtual Environ.* 7, 3 (June 1998), 225–240. https://doi.org/10.1162/105474698565686

[113] Jacob O. Wobbrock and Julie A. Kientz. 2016. Research contributions in human-computer interaction. *Interactions* 23, 3 (April 2016), 38–44. https://doi.org/10.1145/2907069

[114] Andre Zenner and Antonio Kruger. 2019. Estimating Detection Thresholds for Desktop-Scale Hand Redirection in Virtual Reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, Osaka, Japan, 47–55. https://doi.org/10.1109/VR.2019.8798143

A APPENDIX: EDUCATIONAL MATERIAL

As part of this publication, we will provide educational material as supplemental material. It will be distributed under the Creative Commons BY 4.0 license (https://creativecommons.org/licenses/by/4.0). It can then be shared, adapted, and printed as long as the original publication and authors are appropriately cited, and any modifications are indicated. Figures 8, 9, 10, and 11 preview the files that will be made available upon acceptance of the peer-reviewed article.

We will further share a template of the Correspondence Figure as used in Figures 2, 5, 6, and 7, which compares a reference interaction to an implementation in VR. It can be adapted in other publications and will be available as PNG and PSD (Adobe Photoshop) files.

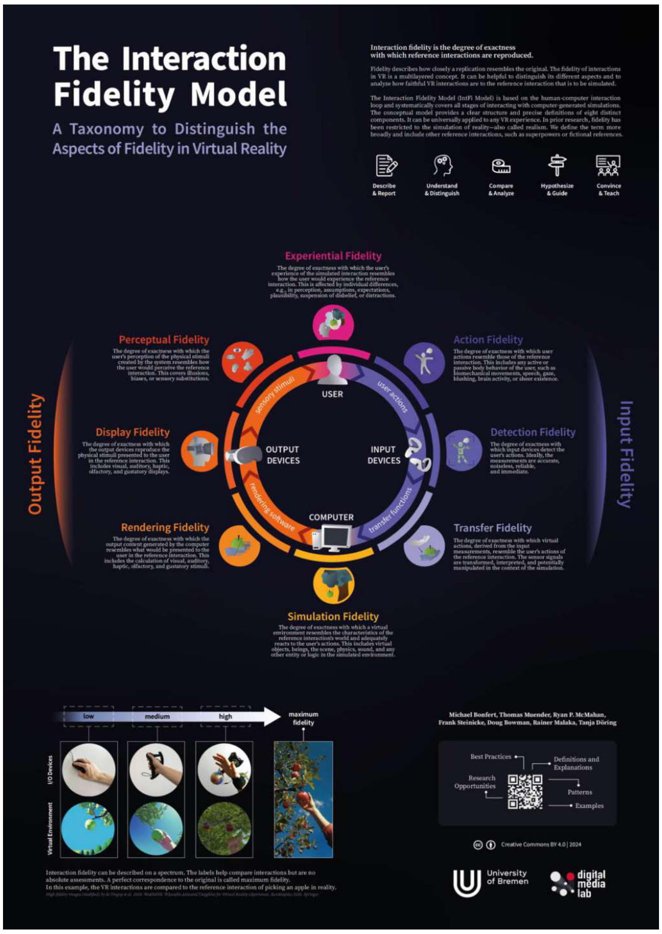


Figure 8: A preview of the large ISO A0 portrait poster. It will be distributed in the highest printing quality after peer reviewing.

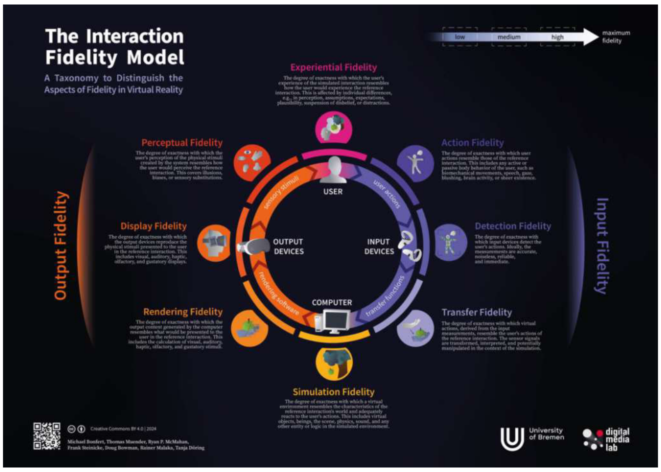


Figure 9: A preview of the ISO A1 landscape poster. It will be distributed in the highest printing quality after peer reviewing.

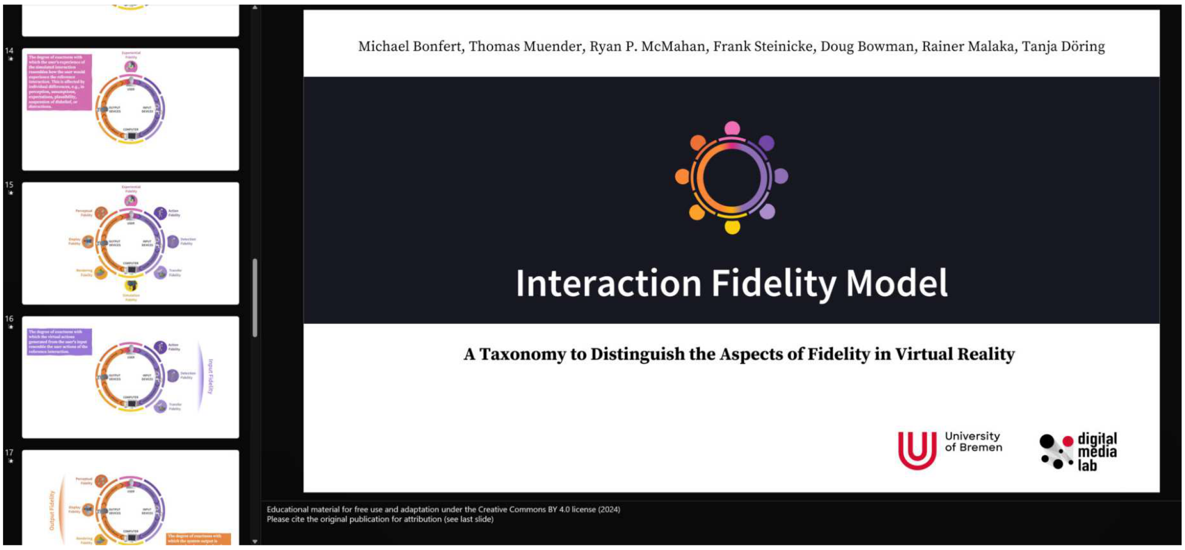


Figure 10: A preview of the slide deck outlining the central concepts of this work. The editable PPTX file will be distributed after peer reviewing.

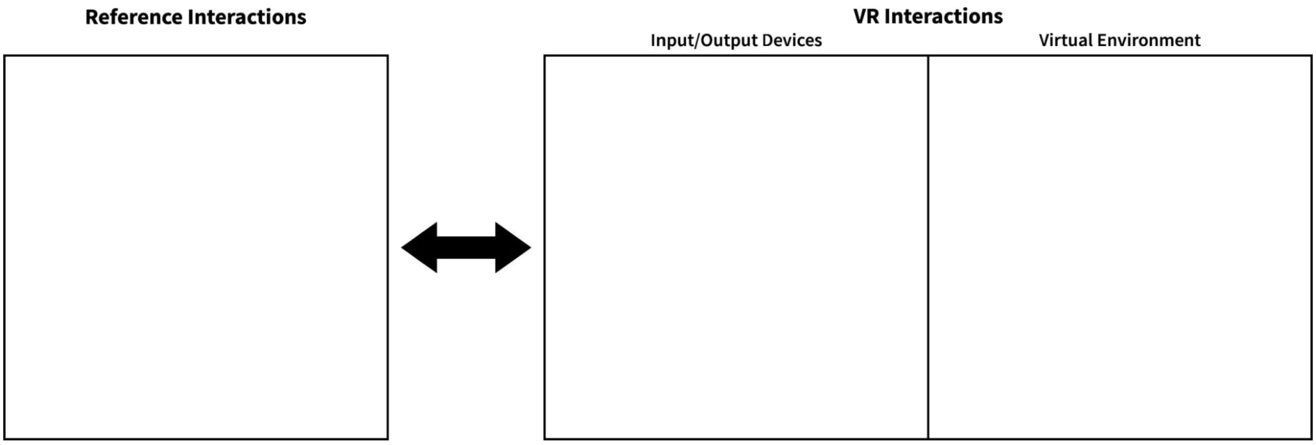


Figure 11: The template of the Correspondence Figure in which other systems can be inserted. The high-resolution PNG and an editable PSD file will be distributed after peer reviewing.

Publication P2

Haptic Fidelity Framework: Defining the Factors of Realistic Haptic Feedback for Virtual Reality

Thomas Muender, Michael Bonfert, Anke Reinschluessel, Rainer Malaka, and Tanja Döring

This publication introduces a framework to determine the haptic fidelity of a VR system. The Haptic Fidelity dimension comprises 14 criteria describing foundational and limiting factors grouped into Sensing, Hardware, and Software. A second Versatility dimension captures the current trade-off between high-fidelity but application-specific and more low-fidelity but widely applicable feedback. The framework was validated by comparing the Haptic Fidelity scores to the measured perceived realism of user studies from 38 papers, which showed a strong correlation.

My contribution: I contributed to the research idea, the theoretical considerations in terms of structure and definitions, the illustrative examples, and the conception of the scoring approach (*conceptualization*). I further contributed to the literature review, the validation method and process (*data curation*), and the writing and editing of the manuscript (*writing – draft, review & editing*).

Published in the *Proceedings of the 2022 ACM Conference on Human Factors in Computing Systems (CHI '22)*

Haptic Fidelity Framework: Defining the Factors of Realistic Haptic Feedback for Virtual Reality

Thomas Muender
thom@uni-bremen.de
Digital Media Lab, University of Bremen
Bremen, Germany

Michael Bonfert
bonfert@uni-bremen.de
Digital Media Lab, University of Bremen
Bremen, Germany

Anke V. Reinschluessel
areinsch@uni-bremen.de
Digital Media Lab, University of Bremen
Bremen, Germany

Rainer Malaka
malaka@tzi.de
Digital Media Lab, University of Bremen
Bremen, Germany

Tanja Döring
tanja.doering@uni-bremen.de
Digital Media Lab, University of Bremen
Bremen, Germany

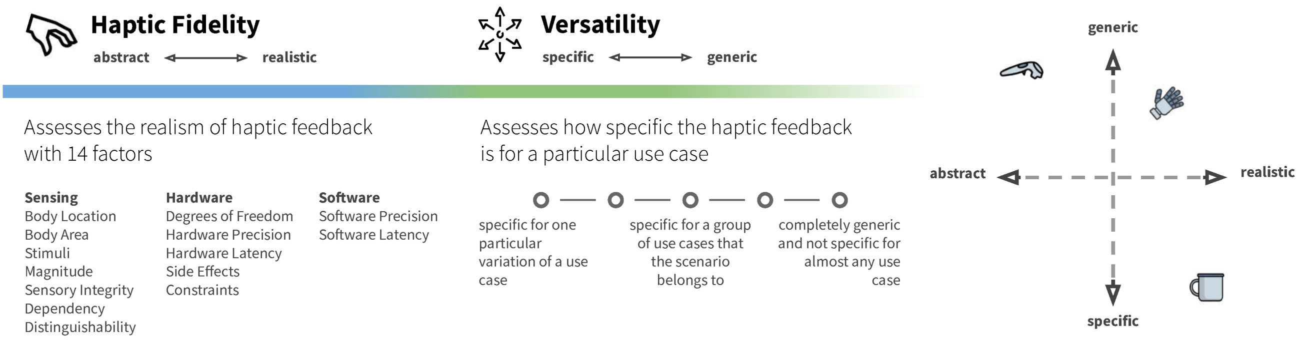


Figure 1: Overview of the two dimensions of the Haptic Fidelity Framework and classification of three examples in the dimensions of the framework for providing haptic feedback for a virtual mug: a standard VR controller with vibration, a force-feedback glove, and a real mug as passive haptic prop.

ABSTRACT

Providing haptic feedback in virtual reality to make the experience more realistic has become a strong focus of research in recent years. The resulting haptic feedback systems differ greatly in their technologies, feedback possibilities, and overall realism making it challenging to compare different systems. We propose the Haptic Fidelity Framework providing the means to describe, understand and compare haptic feedback systems. The framework locates a system in the spectrum of providing realistic or abstract haptic feedback using the Haptic Fidelity dimension. It comprises 14 criteria that either describe foundational or limiting factors. A second Versatility dimension captures the current trade-off between highly realistic but application-specific and more abstract but widely applicable feedback. To validate the framework, we compared the Haptic Fidelity score to the perceived feedback realism of evaluations from 38 papers and found a strong correlation suggesting the framework accurately describes the realism of haptic feedback.

CHI '22, April 29-May 5, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-9157-3/22/04.
https://doi.org/10.1145/3491102.3501953

CCS CONCEPTS

• **Human-centered computing** → **HCI theory, concepts and models**; **Haptic devices**; • **Hardware** → **Haptic devices**; • **Computing methodologies** → *Virtual reality*.

KEYWORDS

framework, haptic feedback, haptics, feedback, virtual environment, immersion, realism, user experience, fidelity, versatility

ACM Reference Format:
Thomas Muender, Michael Bonfert, Anke V. Reinschluessel, Rainer Malaka, and Tanja Döring. 2022. Haptic Fidelity Framework: Defining the Factors of Realistic Haptic Feedback for Virtual Reality. In *CHI Conference on Human Factors in Computing Systems (CHI '22), April 29-May 5, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3491102.3501953

1 INTRODUCTION

Virtual reality (VR) aims to provide compelling experiences for users through immersive technology. Extending this experience with haptic feedback has become a strong focus in research in recent years, intending to create more realistic and engaging virtual environments (VE). One of the big challenges in this field is to create realistic and convincing feedback to simulate the real world or create experiences beyond what is possible in reality. As a result, a large variety of haptic feedback systems have been proposed ranging from standard VR controllers with vibration feedback [57] over custom build controllers simulating shape [74], weight [86], and forces [85] to electric muscle stimulation (EMS) [47], actuated robot arms [58], drones [26], and passive haptic props [53]. These systems differ considerably regarding their technologies, their goals, and approaches on how and where to provide feedback. This variety makes it challenging to compare systems, find comprehensive results in evaluations and inform the design of new systems. Researchers and designers of haptic feedback systems benefit from systematic information on factors influencing the perception of haptics and how the system design relates to these factors. Providing a structured framework that incorporates the underlying mechanics gives them the opportunity to analyze feedback systems and reveal potential for new designs. Current frameworks either focus on the user experience [38], target specific types of feedback, e.g., vibration feedback [54], or propose physical measures that only apply to certain types of feedback systems [59]. To the best of our knowledge, there is no comprehensive framework that incorporates the foundations of haptic perception and is suitable for the whole variety of haptic feedback systems in virtual reality.

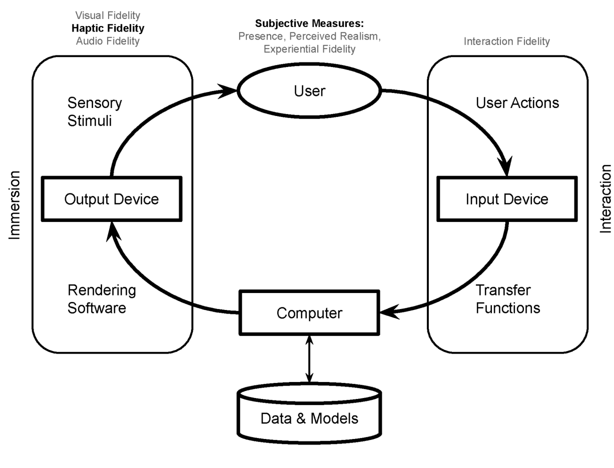


Figure 2: Human-VE interaction loop (first introduced by Bowman and McMahan [9]) with examples for objective measures: Visual Fidelity [9], Haptic Fidelity (presented in this paper), Audio Fidelity, Interaction Fidelity [51] and examples for subjective measures: Experiential Fidelity [2, 44], Presence and Realism [63, 69, 80].

Haptic feedback for VR is created as a response to an interaction with the system. A software then calculates the feedback controlling an output device which acts on the human body and creates sensory stimuli as shown in Figure 2. The user integrates all sensory stimuli into a comprehensive perception of the (virtual or real) world. The resulting user experience is highly subjective and depends on the context, the user's state of mind and past experiences [42]. While haptic feedback is strongly connected to the interaction and the user experience, it is primarily dependent on the output technology and human perception. Therefore it is connected to the concept of immersion describing the potential of a system to enable an elevated user experiences based on the system's rendering software and display technology. It refers to the objective level of sensory fidelity a VR system provides [68]. Other frameworks already rely on the definition of immersion as an objective measure to assess the fidelity of VR systems, such as the visual fidelity [9]. The fidelity of a system describes the quality of being accurate in its output. For haptic feedback "accurate" refers to the capability to produce realistic feedback which is consistent with what users have learned and thus expect from the real world. To assess haptic feedback systems regarding their potential to create realistic experiences, the objective level of sensory fidelity for the haptics must be considered.

In this paper we present the Haptic Fidelity Framework which can be used to assess haptic feedback systems in two dimensions: *Haptic Fidelity* describing the sensory fidelity and the resulting capability to produce realistic or abstract feedback relative to the intent of the system, and *Versatility* representing how specific the haptic feedback is for the particular use case. We validate our framework by applying it to 154 haptic feedback conditions of 38 research papers and compare the *Haptic Fidelity* score to the users' perceived realism. We contribute a comprehensive framework to qualitatively assess all types of haptic feedback systems for VR. It provides researchers and designers of haptic feedback systems with the ability to **anticipate** the fidelity and versatility of a system's feedback, **hypothesize** evaluation results, **compare** different systems or their variations, **decide** which solution best fits a use case, and **understand** the influencing factors of haptic feedback. The framework enables a deeper understanding of haptic feedback and its underlying concepts that can go beyond the application in the context of VR. We provide detailed examples on how to apply the framework in analyzing systems as well as five applications of how researchers and designers can employ the framework in their work with haptic feedback systems. By validating the framework with a wide variety of existing feedback modalities from research, we provide evidence for its usefulness in research and design.

2 RELATED WORK

Haptic feedback has been shown to have an added value for the user experience in general [48] and for immersive VEs specifically [28, 34]. As a result, researchers have developed frameworks to analyze and structure haptic feedback systems. Seifi et al. [65] provide a taxonomy to structure haptic feedback devices based on physical and utility device attributes. Haptic devices were also organized by attributes, categories, and similarities utilizing experts' mental models [66].

Several works propose performance metrics based on physical measures for the design [36] and evaluation of haptic feedback systems [23, 59]. Samur [59] proposes performance metrics based on physical measures for unpowered, powered, and controlled systems (e.g., kinematics, actuation, or impedance) as well as a test bed consisting of six experiments for psycho physical evaluations of haptic feedback systems. While physical metrics such as motion range, peak force, or inertia [23] provide objective measures, each of them is only applicable to a certain type or category of feedback devices and does not provide well defined measures for the whole spectrum of haptic feedback devices.

The user experience of haptic feedback is covered by several frameworks. Hamam et al. [21] introduce a model to capture the quality of experience that integrates both quality of service metrics, e.g., response time and user experience measures. Kim and Schneider [38] define the term "haptic experience" to capture the unique impact of haptic feedback on user experience with a focus on vibrotactile feedback. They propose design parameters, usability requirements and experiential dimensions as well as an initial scale to measure vibrotactile experiences [61]. Furthermore, the language to describe and communicate about haptic feedback has been examined. In a study by Obrist et al. [54], 14 categories of experiential vocabulary have been identified to describe the user experience of vibrotactile feedback. Targeting a wider spectrum of haptic feedback modalities, Schneider et al. [62] identified three themes: the multisensory nature of haptic experiences, a map of the collaborative ecosystem, and the cultural context of haptics for the design of haptic experiences. The survey from Bouzbib et al. [8] investigates the close relationship of interactions and haptic feedback for a wide variety of feedback solutions and propose two dimensions: the solution's degree of physicality and degree of actuation.

While the goal of an improved user experience by adding haptic feedback is clear, the effects thereof are more complicated. Berger et al. [4] investigated the uncanny valley of haptics showing that haptic feedback can reduce subjective realism if is incongruent with other sensory stimuli. Measuring the effects of haptic feedback on user experience with standardized questionnaires is still challenging. Although some questionnaires integrate optional questions related to haptic feedback, e.g., the Presence Questionnaire by Witmer and Singer [80], or the haptics addition [6] for the User Experience Questionnaire (UEQ) [41], they often do not cover the variety of haptic feedback methods and only cover specific aspects of the user experience.

In contrast to the focus on the subjective user experience, the effects of immersion as an objective measure for the technology [68] enabling a heightened user experience have also shown to be valuable for the analysis of VR applications. Bowman and McMahan [9] analyze the effect of visual fidelity and its hardware and software factors, e.g., field of view, display resolution, and frame rate, relating them to immersion benefits and the application effectiveness in creating a realistic and believable VE. McMahan et al. [50] have also evaluated the effects of display fidelity and interaction fidelity [51] and suggest a framework to analyze the realism and naturalness of interaction. They found that both display and interaction fidelity significantly affect performance and subjective measures of presence, engagement, and usability. Furthermore, Lindquist et al. [45] studied the effect of audio fidelity in VR indicating that ambient sound and sound realism increase perceived realism. In a meta-analysis, Cummings and Bailenson [13] showed that factors of technological immersion have a medium-sized effect on presence and that aspects of user-tracking, stereoscopic visuals, and wider fields of view are significantly more impactful than others.

The existing literature shows that several frameworks have been proposed focusing on physical measures or the user experience of VEs with haptic feedback. While immersion and its objective measures have been shown to give valuable insight to the underlying effects of user experience and constitute a predictor of presence or perceived realism, to the best of our knowledge there is no comprehensive framework integrating perceptual and technological measures of immersion covering the whole variety of haptic feedback systems.

3 HAPTIC PERCEPTION

Humans perceive haptics through their cutaneous and kinesthetic systems enabling the perception of material characteristics of surfaces and objects as well as position and movement of their own body. Each of these systems relies on various receptors, which are distributed across the skin surface for the cutaneous system, and the muscles and the tendons for the kinesthetic system.

The cutaneous system is concerned with the perception of tactile information based on four different kinds of mechanoreceptors and general free nerve endings. They provide information about skin stretch and sustained pressure (Ruffini corpuscles), pressure changes and vibrations (Pacinian corpuscles), shape and texture changes (Meissner's corpuscles), pressure, position, and deep static touch features (Merkel nerve endings), as well as touch, pressure, and stretch (free nerve endings) [18, 43]. In addition, thermoreceptors provide information about heat and cold. The receptor density differs on the human body which leads to variations in the spatial resolution of tactile perception. The density was evaluated through two-point touch distance and point localization threshold experiments [43].

The kinesthetic perception originates from mechanoreceptors in the muscles and joints, which provide information about changes in muscle length and velocity (primary and secondary muscle spindles), muscle tension (Golgi tendon organs) and joint extension or flexion (joint receptors) [32, 43]. These signals build our awareness of where our limbs are in space, how they move, and of mechanical properties of objects (e.g., weight, compliance).

All cutaneous and kinesthetic inputs are combined and weighted into a comprehensive haptic perception of the world. The combination of these stimuli enables the perception of haptic features, like the volume and global shape of larger objects, which go beyond low-level tactile properties.

4 THE FRAMEWORK

In this work, we develop a comprehensive framework that incorporates the full spectrum of haptic feedback systems independent of the used technology or addressed body parts. Our goal was to identify a set of criteria that systematically assesses haptic feedback systems. While the user experience of such systems is extremely important for VR applications, it is also highly subjective making it challenging to find a consistent understanding of the underlying processes. Therefore, we aim for more objective criteria that rely on the foundations of perceptual psychology and properties of the technology. As these criteria represent measures of immersion, they represent indicators that can influence the user experience (e.g., perceived realism) and should be considered when evaluating haptic systems. While physical measures also provide objective criteria, they are only applicable to certain feedback devices, e.g., force is only a valid measure for active devices with a motor but not for passive haptic props. In order to cover the full spectrum of possible feedback devices we integrate a layer of abstraction by assessing the match to reality for the particular use case of a system. The framework consists of two dimensions. The *Haptic Fidelity* dimension provides 14 criteria for assessing how abstract or realistic a system can produce haptic stimuli. A second *Versatility* dimension describes how specific the haptic feedback is for the particular use case. The two dimensions cover the trade-off between very realistic feedback for a specific scenario, and creating feedback that can be applied to many different scenarios but is more abstract. The target users of our framework are researchers, developers, and designers. It provides them with means to anticipate, hypothesize, compare, decide, and understand haptic feedback devices and their fidelity as well as versatility.

4.1 Method

For the development of the framework, we used an iterative process over the course of several months to define the framework and identify its factors. We started with an inductive approach by collecting relevant factors that influence the quality of haptic feedback. In addition, we deductively identified three important categories based on the definition of the human-VE interaction loop: *Human Sensing*, *(Display) Hardware*, and *Rendering Software*. To find relevant factors for the *Sensing* category, we conducted a comprehensive literature review on how humans sense haptics, presented in section 3, and studied findings from perceptual psychology [18, 32, 35, 43]. For the *Hardware* and *Software* categories we examined common technology metrics for haptic devices [23] and general output devices [50] that would be relevant for rendering haptic feedback. In the iterative process, the team of authors continuously proposed new or adapted existing factors and then checked if they can accurately represent a variety of haptic feedback systems. We selected nine papers [5, 14, 15, 19, 46, 60, 83, 84, 86] with diverse haptic feedback modalities as references for repeated evaluations of the factors. Our assessment and adaption of factors considered both the conformance to concrete systems and the general applicability to a wide variety of different systems. The evaluation of factors was conducted with a critical mindset to challenge if they accurately and fully cover the whole range of feedback systems. After several iterations, the process resulted in a set of 14 factors.

We then conducted a workshop with seven experts in VR and haptic feedback research. The experts either had designed and built haptic feedback systems themselves or had applied them in their research. Their expertise ranged from material aspects and tangibles (14 years), physical objects in VR and actuated devices (6 years), feedback for medical applications (6 years), interaction techniques in VR (5 years), haptic feedback for psychotherapy in VR (3 years), sports and exergames in VR (3 years) to everyday materials in VR (4 years). During the workshop, the experts were asked to intuitively rate each of the feedback systems from the above-mentioned papers and then to apply the factors to the systems. In a discussion round we asked the experts how well the factors represented the systems, how well the factors could be applied to all systems and if some aspects of the systems were not covered. We received positive feedback for the presented 14 factors, but five experts mentioned that they were missing factors that describe how versatile a system is. They wished for factors that described how generic a feedback system is to be used for different applications. As these aspects do not describe qualities of the haptic feedback but rather how the system can be used, we introduced a second orthogonal dimension to the framework integrating this aspect.

4.2 Haptic Fidelity

In the following, we define *Haptic Fidelity* and introduce the 14 independent factors characterizing this dimension. An overview is shown in Table 1.

4.2.1 Definition. *Haptic Fidelity* describes an objective measure for the qualities regarding the realism of a haptic rendering system. It takes into account how the haptics are rendered and which haptic receptors are addressed but does not describe how a user will experience the haptics. It provides a measure of how realistic the system can reproduce a haptic experience through its rendering mechanisms, thus the potential for a realistic perception from the user. A system with high *Haptic Fidelity* should provide realistic rendering mechanisms that address the same haptic receptors (e.g., skin stretch, pressure, or force) in the same intensity as in the real world. In addition, no noise or confounding factors should interfere with the haptic perception. As the actual haptic quality of the system, which is how a user will experience the system, is not directly linked to how the haptics are rendered, a lower *Haptic Fidelity* is not of inferior quality but displays a different kind of haptic quality. We refer to this quality as "abstract" as it stands in contrast to the realistic rendering mechanisms of a system with high *Haptic Fidelity*. A system with an abstract quality might only address a small number or a single type of haptic receptors. It might also facilitate a particular haptic stimuli to convey the perception of other haptic impressions (e.g., use vibration to simulate a force). In addition, the system might have limiting factors such as imprecision or noise.

4.2.2 Haptic Fidelity Factors. This dimension incorporates 14 independent factors from Table 1 to assess *Haptic Fidelity*. They are divided into three categories: Sensing, Hardware, and Software. The factors are further differentiated between *foundational factors* (F) that describe the features of a system and the value they provide, and *limiting factors* (L) that comprise factors negatively impacting the perception. The set of *foundational factors* (such as *Magnitude* or *Sensory Integrity*) represent the added value of the system. *Limiting factors* on the other hand can merely diminish this value. If a limitation is only minor or does not apply to a system at all, the overall value of the system does not change, while major limitations can drastically impair the whole system. We consider the latency of a system, for example, as a *limiting factor*, because it can only have a negative impact though never enhance a system. Factors that do not apply to every feedback system are also included in the *limiting factors* as they do not impact the value of a system if they are not applicable. Combining all individual factors into one overall

Table 1: Overview of the three categories *Sensing, Hardware* and *Software* with the 14 individual factors of the *Haptic Fidelity* dimension including indication for foundational (F) and limiting (L) factors and brief descriptions.

Haptic Fidelity		
Sensing		
Body Location	F	The degree to which the same location(s) on the body or body parts of the user are involved.
Body Area	F	The degree to which the same extent of body surface of the user is involved.
Stimuli	F	The degree to which the same haptic receptors of the user are involved.
Magnitude	F	The degree to which the same intensity and variation of stimuli are involved.
Sensory Integrity	F	The degree to which the haptic stimuli and stimuli of other modalities match regarding the intent of the system.
Dependency	L	The degree to which the absence of different dependent haptic stimuli that are usually perceived together in reality has an impact on the haptic perception of the system.
Distinguishability	L	The degree to which the distinguishability of different physical properties rendered by the system has an impact on the haptic perception of the system.
Hardware		
Degrees of Freedom	F	The degree to which the system can provide haptic feedback with the same degrees of freedom.
Hardware Precision	F	The degree of detail to which the hardware is able to create the intended haptic feedback.
Hardware Latency	L	The degree to which the hardware latency has an impact on the haptic perception of the system.
Side effects	L	The degree to which the system creates unintended haptic stimuli.
Constraints	L	The degree to which the system constrains the user's movement other than in the intended way.
Software		
Software Precision	F	The degree of detail to which the software is able to simulate the intended haptic feedback.
Software Latency	L	The degree to which the software latency has an impact on the haptic perception of the system.

score, the *Haptic Fidelity* dimension provides a single measure for haptic rendering systems that describes how abstract or realistic the system can potentially provide haptic feedback to a user for a particular use case.

In order to cover the whole variety of haptic feedback devices, the factors of this dimension rely on objective measures but are assessed on a more abstract level. While physical measures can only be applied to certain aspects of haptic feedback, the individual factors represent higher-level concepts that can include multiple physical measures of the same kind, e.g., the *Magnitude* factor can describe the strength of a force, a temperature difference, or the amplitude of a vibration. Further, the use case in which a feedback system is applied plays a central role for the assessment of fidelity. To assess whether a system provides realistic feedback, the same physical measures might be compared, but the order of magnitude could be completely different. A small force to the fingertip might be considered realistic to simulate the touch of a soft material but the same force would not be considered realistic to simulate weightlifting. In addition, the impact of limiting factors might be different depending on the use case. The impact of latency might be higher for simulating force feedback when playing tennis than for simulating the heat from a campfire. Therefore, the factors of the *Haptic Fidelity* dimension can only be evaluated in relation to what is intended to be conveyed by the system. If a system is designed to simulate the feedback of punches from boxing, it should only be evaluated how realistic these punches can be represented but not how realistic the system can, for example, represent touching flowers. Therefore, *Haptic Fidelity* and its factors are relative scales that capture the qualitative assessment of how realistic or abstract the feedback is for a particular use case.

In the following, the individual factors are described in detail. The descriptions contain the phrase "The degree to which the same ..." where "the same" refers to how or where the intended haptics would be perceived in reality. Each foundational factor is rated between "no match at all" and "complete match" with reality while limiting factors are rated between "no impact" to "very strong impact" on the perception. The following factors are part of the **Sensing** category about the human capabilities to sense haptic stimuli.

Body Location (F) *The degree to which the same location(s) on the body or body parts of the user are involved.*
This factor describes where on the user's body the haptic feedback is created by the system and to what extent this is in line with where one would perceive the stimulus in the natural occurrence of the intended haptics. This scale is grounded in the human ability to localize a haptic stimulus on the body (bodily localization [43]).

Example: Using EMS on the upper and lower arm to simulate lifting a virtual box, like the system by Lopes et al. [47] does, would get a medium-high score as it correctly involves the arms to simulate the weight of the box, but does not give feedback on the fingers and hands where the box is touched.

Body Area (F) *The degree to which the same extent of body surface of the user is involved.*
This factor describes how well the system provides haptic feedback to the same extent of area on the user's body where one would get feedback in the natural occurrence of the intended haptics. This scale is grounded in the spatial resolution of haptic receptors in the human body and the ability to relate stimuli to each other, forming a consistent sensory impression [43]. The differences in density of haptic receptors should be considered for this factor. Hence, a system that intends to convey haptics to the fingertips should match the area more precisely than a system affecting the upper arm.

Example: Simulating the haptic feedback of a boxing punch with a small, actuated plate on the forearm of a user, like the Impacto system [46] does, involves a significantly smaller area than a boxing glove would impact. Therefore, such a system would get a medium-low score.

Stimuli (F) *The degree to which the same haptic receptors of the user are involved.*
This factor describes how well the system stimulates the same haptic receptors as they would be stimulated in the natural occurrence of the intended haptics. It is grounded in the existence of different haptic receptors in the human body as outlined in section 2. Each

kind of receptor responds to different forms of stimuli, making it possible to perceive and differentiate a variety of haptic properties [43]. While each kind of receptor responds to specific low-level stimuli, perceptual psychology classified haptic stimuli into higher-level perceptions that can be distinguished by humans. To rate this factor without deeper knowledge in perceptual psychology, we provide Table 2 that relates the perceivable haptic stimuli with physical properties that can be rendered by haptic feedback devices.

Example: Using the vibration of a VR controller to give feedback for the contact force of virtual walls, as it is done in one of the conditions by Boldt et al. [5], involves completely different haptic receptors leading to a low score.

Magnitude (F) *The degree to which the same intensity and variation of stimuli are involved.*
This factor describes how well the system creates the same strength (e.g., same force) or variation (e.g., same texture) of haptic stimuli compared to the natural occurrence of the intended haptics. The scale is grounded in the ability of human haptic receptors to perceive different variations of stimuli. When rating this scale, the just noticeable difference in sensations from a haptic system should be considered [33].

Example: Simulating the haptic feedback of a boxing punch, like the Impacto system [46] does with EMS and a small, actuated plate on the forearm, can only represent the force to a medium degree and the pressure applied to the forearm to a low degree. Therefore, such a system would get a medium-low score.

Sensory Integrity (F) *The degree to which the haptic stimuli and stimuli of other modalities match regarding the intent of the system.*
This factor describes to what extent the haptic stimuli match with the perception of the other senses that are also addressed by the system and if this match is in line with the intent. It is based on the human ability to integrate all senses into a consistent perception of the world. The senses are weighted differently for this integration; especially vision has been found to be weighted more strongly [43]. This visual dominance effect [24] leads to the possibility that the visual perception can influence how the haptics are perceived.

Example: A physical sandbox with a VR visualization, e.g., as by Fröhlich et al. [15], can display water evoking the expectation that the haptic perception will be consistent with it and feel like water, but the user will only perceive the haptics of sand. This can be rated with a medium degree of integrity as it does not match to full extent but is also better than perceiving a solid surface or nothing at all.

Dependency (L) *The degree to which the absence of different dependent haptic stimuli that are usually perceived together in reality has an impact on the haptic perception of the system.*
This factor describes if the system creates different haptic stimuli that are usually perceived together in nature, e.g., weight together with weight distribution when lifting an object. The fact that these dependent haptic stimuli are naturally perceived together create the expectation to always perceive dependent haptic stimuli together. Therefore, this factor is based on the integration of different haptic stimuli and the learned expectations which haptic stimuli are generally perceived together. This factor is a limiting factor because not all haptic stimuli have other dependent stimuli.

Example: A force feedback glove can be used to render the weight, contact force, volume and shape of objects that can be touched and lifted. But it does not provide dependent haptic stimuli like the texture or temperature of the object, which would be perceivable when touching an object in reality. Therefore, it has some limitations and a medium-low impact on the haptic perception can be assumed.

Distinguishability (L) *The degree to which the distinguishability of different physical properties rendered by the system has an impact on the haptic perception of the system.*
This factor describes if different physical properties that are intended to be conveyed by the system can be distinguished by either targeting different haptic receptors shown in Table 2, or through spatial or temporal separation. Different haptic receptors and the integration of different haptic stimuli makes it possible to distinguish a variety of haptic properties. While in nature each object has individual haptic properties, haptic feedback systems are sometimes only capable of providing a limited number of haptic stimuli. Remapping haptic stimuli to represent other physical properties is often used in these cases, e.g., using vibration to represent the contact force of an object. This might work well with a distinct mapping where users can clearly identify and learn the remapping, but can lead to confusion and unrealistic sensations from ambiguous mappings. This factor is a limiting factor because not all systems intend to render multiple physical properties or use any kind of remapping.

Example: A system that uses vibration to render texture and contact force of an object when touching it would make it challenging to distinguish which of the two physical properties is currently rendered or varied. Therefore, such a system would get a high rating due to its limitations.

The following factors are part of the **Hardware** category about the devices that are used to create the haptic feedback.

Degrees of Freedom (F) *The degree to which the system can provide haptic feedback with the same degrees of freedom.*
This factor describes if the hardware of the system provides at least the same number of degrees of freedom (DoF) as in the natural occurrence of the intended haptics. The system can have more degrees of freedom than they require making the system more versatile. Naturally, objects that can be freely moved and rotated have 6 DoF while slide doors would only have 1 DoF and walls none. This factor is based on the fact that haptics are naturally present in multiple dimensions which must be represented by technical solutions. If the system provides a certain number of DoF but the user's range of motion is limited within these through the system, it is not considered in this factor but part of the factor *Constraints* below.

Example: The Aero-Plane system [31] is a custom controller with two vertical propellers meant to simulate the forces of a ball rolling on a plane, or the motion of food in a pan. The system offers two DoF through the actuation of the propellers creating forces in the left/right and up/down directions. To properly simulate the intended scenarios, a third DoF would be necessary to also create forces in the front/back direction. Therefore, the system receives a medium-high score.

Hardware Precision (F) *The degree of detail to which the hardware is able to create the intended haptic feedback.*
This factor describes to what extent the system can reproduce the detail of haptic feedback compared to the natural occurrence of

Table 2: Mapping of perceivable haptic stimuli (listed vertically on the left) and physical properties (listed horizontally above) of objects and surfaces. The haptic stimuli are based on perceptual psychology literature [43] classifying low-level stimuli into higher-level perceptions that can be distinguished by humans. We mapped the relationships of these haptic stimuli to physical properties that might be rendered by haptic feedback systems based on the involvement of the different haptic receptors. Forces are classified as either body forces (generated through the interaction of physical bodies with mass, force, velocity, etc. or force fields such as gravity) or contact forces (produced by direct physical contact) [52]. We distinguish four categories of involvement:
■ (green) This haptic stimulus is always involved in the perception of this physical property;
■ (orange) This haptic stimulus is always involved in the perception of this physical property when the object or body is in motion;
■ (yellow) This haptic stimulus is possibly involved in the perception of this physical property depending on how the object or surface is touched;
□ (empty) This haptic stimulus is normally not involved in the perception of this physical property.

physical properties / haptic stimuli	weight (distribution)	volume	shape	texture	compliance	friction	temperature	contact force	body force
skin stretch	■ (yellow)			■ (yellow)	■ (yellow)	■ (green)		■ (yellow)	■ (yellow)
sustained pressure	■ (green)	■ (green)		■ (yellow)		■ (yellow)		■ (yellow)	■ (yellow)
pressure changes	■ (orange)		■ (orange)	■ (yellow)	■ (green)	■ (yellow)			■ (orange)
vibration				■ (yellow)		■ (yellow)			
shapes / edges		■ (yellow)	■ (green)	■ (yellow)					
texture				■ (green)					
temperature							■ (green)		
proprioception		■ (green)	■ (yellow)		■ (yellow)			■ (yellow)	■ (yellow)
constant force	■ (yellow)	■ (green)				■ (yellow)		■ (green)	■ (yellow)
force changes	■ (orange)		■ (orange)		■ (green)	■ (yellow)			■ (orange)

the intended haptics. It is based on the necessity of mechanical components with varying and unavoidably finite accuracy. This can include the resolution of building blocks or the detail of object resemblance. Related aspects that can influence the precision, e.g., calibration imprecision or joint tolerance, should also be considered for this factor.

Example: The precision of a pin-based system to render the shape of objects is dependent on the number and size of the individual pins. The PoCoPo system [84] uses a small number of 1 cm wide pins which is a low resolution compared to the sensory resolution of the hand. It would therefore get a medium-low score. A study by Muender et al. [53] compares three types of passive haptic props with different degrees of detail. The differences in prop detail can be expressed in this factor.

Hardware Latency (L) *The degree to which the hardware latency has an impact on the haptic perception of the system.*
This factor describes the influence of the hardware latency on the haptic perception. It is based on the fact that technical solutions often require time to transmit signals and mechanically change their states. While in nature haptics are almost always perceived immediately to an action, the latency of haptic feedback systems can lead to a delay in the haptic perception and therefore result in unrealistic sensations. The time until a delay is noticed and the influence of delay on the perception depends on the kind of stimuli and situation [30]. Delays over 25 ms to 100 ms are commonly reported to be noticeable and have an impact on performance. This is a limiting factor because latency can only have a negative impact on the haptic perception and will never enhance a system.

Example: The Shifty system [86] actuates a weight up and down in a tube to simulate different weight distributions, which takes up to 2.8 s. Since this latency is not constantly noticeable but only when a different virtual object is picked up, and since it is mitigated by adjusting the weight already before grasping, the system would get a medium score.

Side Effects (L) *The degree to which the system creates unintended haptic stimuli.*
This factor describes how the haptic perception is influenced by any side effects that the hardware of the system creates. This can include the vibration of motors, pain from EMS, or forces from moving parts. It is based on the fact that technical devices can produce unintended haptic stimuli which may lead to distraction, irritation, and unrealistic sensations. This is a limiting factor because side effects can only have a negative impact on the haptic perception and never enhance a system.

Example: EMS systems, such as by Lopes et al. [47], can cause unpleasant traction, tickle, or even pain leading to a medium score. Other systems that have actuated parts, as the Shifty system [86], will create unintended counter forces and inertia from the device adjusting, leading to a low score.

Constraints (L) *The degree to which the system constrains the user's movement other than in the intended way.*
This factor describes how the haptic perception is influenced by constraints that the hardware imposes on the user's range of motion. This could be caused by joint limits, wire length or colliding parts of the hardware. It does not consider if the system intentionally limits the user's range of motion to render haptics. This factor is based on the fact that technical solutions might restrict the user's freedom of movement because of the system design or mechanical components such as wearables or wires. This is a limiting factor

because constraints can only have a negative impact on the haptic perception and will never enhance a system.

Example: An exoskeleton or robotic arm that limits the range of motion of the user's arms due to its joint limits or world-grounding, as it is the case in [58], would get a medium score.

The following factors are part of the **Software** category about the software that is used to create the haptic feedback.

Software Precision (F) *The degree of detail to which the software is able to simulate the intended haptic feedback.*
This factor describes how accurately the software can calculate the haptic feedback to be rendered. This can include how accurate colliders represent the shape of an objects or how accurate a physical simulation can calculate the compliance of an object. It is based on the fact that the feedback has to be calculated by software before it can be displayed by hardware.

Example: If an exoskeleton is used to render the contact force of an object but the collider to represent the object uses an inaccurate low-resolution mesh, the hardware can at most render the haptics with this low resolution. Depending on how low this resolution is, the system would have a low or medium-low score.

Software Latency (L) *The degree to which the software latency has an impact on the haptic perception of the system.*
This factor describes the influence that the software latency has on the haptic perception. It is based on the fact that the necessary calculations take time and may cause a noticeable delay. The same threshold for noticeable delays as in *Hardware Latency* apply for this factor. This is a limiting factor because the presence of latency can only have a negative impact on the haptic perception and will never enhance a system.

Example: A system that takes 500 ms to simulate the compliance of an object creates a noticeable delay. A latency of 500 ms can be interpreted as a medium impact on the haptic perception and therefore would get a medium score.

4.2.3 Scoring. The individual factors of the *Haptic Fidelity* dimension can give designers and researchers insight into the importance of certain aspects of haptic feedback in VR or the origin of limitations. Yet, to structure the qualitative assessments of individual factors and effectively compare different systems that have the same or similar use cases, a single comprising score representing the complete *Haptic Fidelity* of a system is required.

$$\underbrace{\frac{\sum_0^{N_F} X_F}{N_F}}_{\text{mean of foundational factors}} \cdot \underbrace{\mathrm{e}^{-0.0027\cdot(\overbrace{\sum_0^{N_L} X_L^2}^{}_{\text{sum of squared limiting factors}})^2}}_{\text{exponential function for the influence of limiting factors on the value of a system}}$$

Figure 3: Equation to calculate a single score for *Haptic Fidelity*, with N_F for the number of foundational factors, X_F for the respective rating of the foundational factors, N_L for the number of limiting factors and X_L for the respective rating of the limiting factor.

To build an overall score, all factors first have to be evaluated in terms of actual numbers. We propose to evaluate each factor on a 5-point Likert-scale (0–4), as it offers a good trade-off between having enough options to differentiate systems[1] but still having distinct differences between ratings to enable clear decisions which rating represents a system best. To calculate a single score for each dimension that properly represents the systems on the respective scale the foundational factors (F) will be averaged as they all represent an added value to the system. Limiting factors (L) on the other hand represent aspects that negatively impact the system and consequently have to lower the overall score. Therefore, the scales for limiting factors must be inverted (0 – no limitation; 4 – strong limitation). In addition, a high limiting factor can drastically influence the value of a system, e.g., if the software latency is 5 seconds, it does not matter how fast or accurate the hardware is or if the right receptors are targeted, the latency still limits the whole systems. To reflect this in the score we propose to square the ratings of limiting factors, giving them increased influence on the final score. To combine the scores of foundational and limiting factors our goal is to calculate a value between 0 and 1 that can be multiplied with the foundational score where 1 implies no limitations leaving the foundational score as is and 0 representing strong limitation completely nullifying the value of the system. To develop a suitable function that results in a fitting overall score, we used an iterative approach where we tested and adapted several different functions and aimed for a good match with the assessments of the experts from the workshop. This approach resulted in an exponential function of the form $e^{-\alpha * x^2}$ to transform the sum of all squared limiting factors into the desired range. This function satisfies two requirements: First, $f(0) = 1$ with 0 being the sum of all limiting factors representing no limitation at all. Second, the fact that one high limiting factor can already drastically lower the value of a system is represented in the exponential function having a quick drop off. For multiple high limiting factors, the multiplication factor approaches zero representing the strong negative impact. To determine the α-factor for the exponential function we use the rational that if one limiting factor receives a maximum score (4) the foundational score should be reduced by half, which was based on a good fit to the assessments of the experts. As a result we use the formula shown in Figure 3 to calculate the final score for *Haptic Fidelity*.

4.3 Versatility

The *Versatility* dimension of the framework describes a measure on how specific a system is in providing haptic feedback for a particular application. A similar definition of versatility as an important attribute of haptic systems was given by Seifi et al. [66]. This dimension exists orthogonal to the *Haptic Fidelity* dimension as it represents the trade-off between highly realistic but application-specific and more abstract but widely applicable feedback. Systems that provide abstract feedback are naturally more widely applicable for different application scenarios as the feedback can be repurposed to represent any kind of other more specific feedback (e.g., vibration feedback is used to represent contact force or weight).

[1] Three options as used for the interaction fidelity [49] do not properly capture the differences of certain factors like the *Body Area*.

Realistic feedback, on the other hand, can be achieved by designing systems that are custom-made for a particular application. Only addressing receptors at the exact body position necessary for the intended application makes these systems extremely specific to this scenario. Placing feedback systems in the space of these two dimensions should give researchers and designers a better understanding on how the realism of feedback is connected to the possible use cases and how more abstract feedback can be applied to different applications. While *Haptic Fidelity* was assessed by individual factors, the *Versatility* dimension provides a single factor rating haptic feedback systems on a scale from specific to generic. How a system is rated on this scale does not have any implication for the overall quality of a system but rather represents how versatile it can be used.

We propose to assess *Versatility* on a 5-point Likert-scale (0–4) with 0 representing systems that are extremely specific to the intended application and 4 representing systems that are generic in its feedback. As the specificity of a system is not based on objective measures to the same degree as the factors from *Haptic Fidelity* are, we provide categories describing the five levels of this scale. The examples provided for the following categories are illustrated in Figure 4.

- 0 The systems feedback is specific for one particular variation of a use case (e.g., for a particular climbing wall, where other variations of haptics cannot be rendered [64]).
- 1 The systems feedback is specific for a particular use case (e.g., boxing [46]) and supports different variations (e.g., the system could provide feedback for other variations like different boxing fights or even other kinds of martial arts).
- 2 The systems feedback is quite specific for a group of use cases to which the particular application belongs to (e.g., all scenarios with small handheld objects [39]).
- 3 The systems feedback is unspecific for a particular use case and generic to a larger group of applications (e.g., a force feedback glove [70] or an exoskeleton).
- 4 The systems feedback is completely generic and not specific for almost any application (e.g., vibration from VR controller as feedback [57]).

These categories provide the means to congruently rate the versatility of haptic feedback systems and generate a score for the *Versatility* dimension. Similar to the *Haptic Fidelity* this dimension has to be rated relative to the intended application of the haptic feedback. This is necessary to generate valid ratings that are consistent between the two dimensions. A standard VR controller with vibration would normally be considered very generic in its feedback. But if the scenario is to render the haptic of a vibrating phone, for example, it has to be considered as more specific for this particular use case.

4.4 Framework Application

Our framework provides the means for a structured and informed assessment of haptic feedback devices that is based on factors, which are evaluated qualitatively in the context of a specific use case. The factors allow for a detailed assessment that covers all relevant aspects of a system. The framework forms the basis for consistent and comparable assessments as always the same aspects are evaluated and none are missed. It also provides a defined vocabulary for a more precise communication of the analysis and results.

The framework is intended to give qualitative insight into the underlying properties of haptic feedback systems. The individual factors provide the means to think about these underlying concepts and support formalizing them in a quantitative way, i.e., with a score. The score is intended as a tool to structure the inspected qualitative aspects. An assessment of a system should not be made in isolation but should always consider how the same system would be rated when certain aspects were changed or in relation to similar systems that can be applied to the same use case. The individual factors have to be evaluated in relation to the intended use case of the system and therefore provide relative insights. For assessments of different systems or system variations to be comparable, they have to be made on the same basis of assessment framed by the particular use case and by similar systems that are employed as reference.

We intend the framework to be used by individuals or groups of researchers and designers to qualitatively assess feedback systems and make relative comparisons within the same use case or similar ones. In the following, we will present five applications of the framework to illustrate how it can be used. Each application is aimed at a certain user group. The required expertise to apply the framework may vary depending on the application but a general overview of different haptic devices that can be used as reference is advised. A basic knowledge can be sufficient when using the framework to understand and learn about haptics while precise anticipation of evaluation results might require a wider overview of feedback possibilities and an understanding of how haptic feedback is perceived. However, even if not all details are taken into consideration, the framework can still provide valuable insights into the qualities of a system by its structured approach.

4.4.1 How to Use the Framework. The following five applications illustrate how the framework can be used in the work of researchers and designers.

Anticipate: Researchers and designers can apply the framework when designing haptic feedback devices or prototypes to anticipate the realism of feedback and versatility of the system. The framework can be used iteratively from early technical concepts to the final building phase to provide indication on the outcome. The individual factors can further provide guidance on what can be adjusted to result in the desired feedback.

Hypothesize: Researchers can use the framework to formulate reasoned hypotheses for their evaluations of haptic feedback systems and discuss the findings based on the dimensions and factors from this framework.

Compare: Researchers and designers can apply the framework to compare device variants, e.g., [71], or different feedback devices for the same use case, e.g., [31, 85]. The framework can be used to identify similarities and differences between the devices. The factors of the framework can give insight where these differences are, how decisive they are, and what the underlying perceptual or

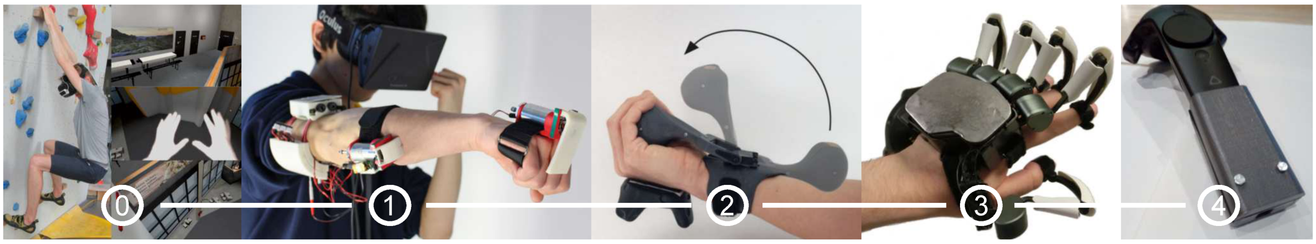


Figure 4: Examples for the five levels of specificity from the *Versatility* dimension, from (0) specific to (4) generic. (0) Schulz et al. [64]: climbing condition; (1) Lopes et al. [46]: EMS and solenoid condition; (2) Kovacs et al. [39]: catching condition; (3) Steed et al. [70]: free condition; (4) Ryge et al. [57]: high fidelity condition. Image copyrights from left to right ©Peter Schulz, ©Pedro Lopes, ©Robert Kovacs, ©Anthony Steed, ©Andreas Ryge.

technical reasons are.

Decide: Product designers can use the framework to make informed decisions which feedback device to select or develop for their specific use case. Researchers can also apply the framework to select appropriate feedback devices to best answer their research questions. The framework can help finding a suitable compromise of overall realism and wide applicability of a product or provide detailed information on the underlying factors to guide decisions.

Understand: The framework can be used to understand influencing factors on haptic feedback. In combination with existing literature, it facilitates a comprehensive understanding of haptic feedback that can inform future research and development. Researchers can use it to soundly interpret the results of evaluations. Further, educators can use it to teach about haptics and give students an overview of existing systems or approaches. Designers can apply the framework to understand why their product performs the way it does. Lastly, the vocabulary established in this framework can help experts express themselves accurately when exchanging with peers.

4.4.2 Example Application of the Framework. In the following, we provide an example describing in detail how different haptic feedback conditions from a research publication can be evaluated with our framework and how the assessments can be used to *Hypothesize* and *Compare* the conditions. In this section, we focus on reporting and motivating the scores for each factor, which are listed in Table 3. Further details on the procedure of applying the framework can be found in subsection 5.2. For comparison, we also provide a classification of ten haptic feedback systems into the two dimensions of the framework in Figure 5.

The paper "The Role of Physical Props in VR Climbing Environments" by Schulz et al. [64] presents a system to simulate the experience of climbing in a climbing gym. The paper has three conditions of which one is real climbing without a VR headset leaving two conditions that can be analyzed. The first condition (C) enables climbing by the means of standard VR controllers with vibration feedback while the second condition (W) enables climbing with hands and feet in VR on a real climbing wall providing passive haptics. The scores for all factors are listed in Table 3.

Table 3: Ratings of all factors for the two conditions of VR climbing by Schulz et al. [64]. C – condition with a standard VR controller; W – condition with a real passive-haptic climbing wall. The 14 factors (either foundational or limiting, see Table 1) result in the score for haptic fidelity (see Figure 3 for the calculation). The score for versatility is expressed on a single scale (see Figure 4).

Factor	**C**	**W**
Body Location	1	4
Body Area	2	4
Stimuli	1	4
Magnitude	0	4
Sensory Integrity	1	4
Dependency	3	0
Distinguishability	0	0
Degrees of Freedom	0	4
Hardware Precision	0	4
Hardware Latency	0	0
Side effects	0	0
Constraints	0	0
Software Precision	4	4
Software Latency	0	0
Haptic Fidelity	**0.9**	**4.0**
Versatility	**4**	**0**

For the *Body Location*, the standard VR controller (C) only provides feedback to the hands but not to the feet, arms, legs, shoulders, or chest that would also experience the forces when climbing. Condition (W), on the other hand, involves all these body parts. The *Body Area* of the hands and feet is involved in climbing on a real climbing wall (W), whereas (C) only involves the skin surface of the hands. The controller (C) provides vibrations as haptic *Stimuli*, which do not naturally occur in climbing, as well as texture, shape, and contact forces but no stimuli for pressure, skin stretch, or forces of the user's body weight (cf. Table 2 for a general overview on different haptic stimuli). The provided stimuli intensity does not match at all with climbing leading to a low score for *Magnitude*. In contrast, the real climbing wall (W) creates the correct *Stimuli* with the right *Magnitude*. Because of this there is also a good match with the visual feedback of the system creating a high value for *Sensory Integrity* (W), whereas the controller (C) only has a limited match with the visuals. The absence of important haptic stimuli for the body weight or pressure on the fingers in the controller condition (C) creates a noticeable limitation in *Dependency*. For both the *Degrees of Freedom* and *Hardware Precision*, the real climbing wall (W) receives high scores as real grab handles and precise tracking are used. The controller (C), on the other hand, cannot represent the intended haptics of climbing to any precision and the 1 DoF of vibration does not match with the required DoF. The *Hardware Latency* is imperceptibly low for both conditions and imposes no limitations. There are no *Side Effects* or *Constraints* in neither of the conditions. The paper does not explicitly report on how the software calculates the haptics, but it is reported that the Unity game engine is used. Thus, it can be assumed that standard colliders with a high *Software Precision* and low *Software Latency* are used. After scores for all 14 factors are identified, overall haptic fidelity scores can be calculated based on the formula given in Figure 3. In our example, this results in a haptic fidelity score of 0.9 for the controller condition (C) and a score of 4.0 for the wall condition (W). Regarding *Versatility*, the climbing wall (W) is extremely specific for its scenario and can also only create feedback for this particular variation of a climbing wall, which is assessed with a score of 0 (category 0 in Figure 4). The controller (C) is considered completely unspecific for the climbing scenario, and thus receives a score of 4 (category 4 in Figure 4). Figure 5 presents the classification of these two conditions on the two dimensions *Haptic Fidelity* and *Versatility* together with eight other research systems for comparison.

The authors of the paper could apply this analysis to support their hypothesis that presence is expected to be higher for users in condition W than in condition C. As perceived realism is one key aspect of presence, the result of more realistic feedback for condition W (4.0) compared to condition C (0.9) supports this hypothesis. Further, the authors could have used this analysis to discuss their results and argue that, for example, the *Magnitude* had a great impact on this result. Comparing the two conditions, it becomes clear that the condition C provides highly versatile feedback while the W condition is focused on realistic feedback for one particular use case.

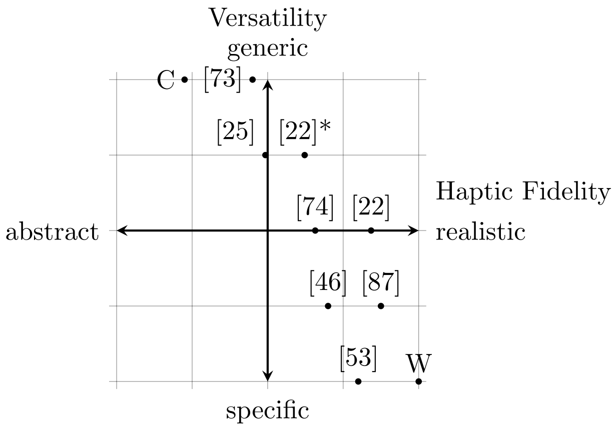


Figure 5: Exemplary classification of haptic feedback systems from research. C and W are the two conditions from VR climbing [64] in Table 3; "Does it feel real?" [53]: Lego condition; HapTwist [87]: Rubik's-Twist-based condition; Impacto [46]: high EMS and high solenoid condition; PuPoP [74]: Quidditch condition; Haptic Around [22]: hybrid condition; Haptic Around [22]*: controller-based condition; DextrES [25]: brake and piezo condition; DualVib [73]: chainsaw force & texture condition.

5 VALIDATION

In order for the framework to be of practical value for researchers and designers, so that they can compare haptic feedback systems and draw conclusions for their research and the design of new feedback methods, it is necessary to demonstrate that the framework is related to the user's actual perceived realism of a system. We formulate the following hypothesis describing the relation between the *Haptic Fidelity* dimension of the framework and the perceived realism to guide our validation procedure.

> **Hypothesis 1**: The haptic feedback of a system that has a high potential to produce realistic feedback, as described by the *Haptic Fidelity* score, will also be perceived as more realistic by users than a system that has a more abstract feedback potential.

The *Versatility* dimension was introduced to represent the potential trade-off between highly realistic but scenario-specific and more abstract but widely applicable feedback. We formulate a second hypothesis describing the resulting relation between the *Haptic Fidelity* and *Versatility*.

> **Hypothesis 2**: In haptic feedback systems, there is a trade-off between its *Haptic Fidelity* and its *Versatility*.

To validate these relationships, it is required to apply our framework to different haptic feedback modalities and evaluate the perceived realism of these modalities. However, it is not practical to evaluate the framework on some experiment as only a few haptic feedback modalities can be compared in a single user study. This does not represent the broad variety of different haptic feedback methods that should be covered by the framework and therefore would not give meaningful insights. Therefore, to validate our framework, we will build on the results of previous research papers which evaluated the perceived realism of different haptic feedback modalities in VR. By applying our framework to the feedback modalities of existing research and correlate the scores to the effect that the feedback modalities had on the perceived realism, we take the first step to demonstrate the validity of our framework.

5.1 Paper Selection

To find relevant papers that can be analyzed with our framework and give insight in the perceived realism, we conducted a literature review searching the ACM Digital Library and IEEE Xplore databases for papers related to the topics of VR and haptics. We selected papers that used VR headsets for the visual presentation and provided some form of haptic feedback. Papers that used other forms of presentation, e.g., augmented reality, 3D displays, or standard displays were excluded as we were especially interested in immersive technology where participants could not see the feedback systems. We further excluded papers published prior to 2000 in order for the technology, especially the VR headsets, to be comparable. As we were interested in user's perceived realism, we did not include technical papers without a user study in our selection. Further, papers in which no real haptic feedback is experienced by users, e.g., pseudo-haptics, or the main goal is not realistic feedback but rather guiding the attention, e.g., alerts, haptics for guidance or professional medical haptic devices were excluded from the selection. From this search we identified 160 Papers fitting our criteria. These papers were then further analyzed regarding whether they evaluated perceived realism in the user study by the means of standard or custom questionnaires. For standardized questionnaires we considered the realism sub-scales of the Witmer and Singer [80], IPQ [63] and SUS Presence [77] questionnaires. Only papers that had at least two conditions that compared different haptic feedback modalities to each other were considered. Most importantly, to ensure that differences in perceived realism measured between the conditions can only be attributed to the differences in haptic feedback, only papers were selected where the haptic feedback modality was the only changing variable between conditions. All other variables, e.g., the visuals and tasks needed to be the same between conditions. We identified 38 papers in our selection of papers that fit these criteria. In total these papers have 154 conditions with varying haptic feedback modalities. The following 38 papers were included in the analysis[2]: [1, 3, 7, 10, 16, 17, 19, 20, 22, 25–27, 31, 37, 39, 40, 46, 47, 53, 55, 57, 58, 64, 67, 70–76, 78, 79, 81, 82, 85–87]

5.2 Process of Applying the Framework

Each of the haptic feedback systems from the 154 conditions were assessed with the framework presented in this paper. A score for each of the 14 factors of the *Haptic Fidelity* dimension was assessed. Based on these, a final *Haptic Fidelity* score was calculated as described in subsubsection 4.2.3. For the *Versatility* dimension, one score was selected according to the five categories of versatility. Applying the framework was done by three of the authors in a discussion round to generate consistent and sound ratings. We chose this form of assessment, as the researchers first had to find a mutual understanding for the haptic feedback systems described in the papers and agree on a consistent basis of assessment for each paper. In an initial test evaluation, we found that quite some discussion already happened during reading the paper and the researchers first had to find a mutual understanding of the system as not all aspects of a system were sufficiently described in some papers. As this discussion already included aspects on how to rate the system, we chose a continuous discussion round for the application of the framework. In order to rate each of the presented factors for all conditions, all researchers introduced arguments for a particular rating for the factor. They discussed the arguments until a consensus was found and one final rating was selected that all would agree on. Papers not always reported all information necessary to rate a factor with absolute certainty. Mostly this was the case for the *Software Precision* and *Software Latency* factors as nothing specific was reported about the software (in most cases only the used game engine was reported). The ratings were then based on the assumption that standard practices were used, e.g., collider to calculate the contact with virtual objects.

We provide all ratings of all factors for the 154 conditions including short arguments on why the researchers chose this rating in the supplementary material.

5.3 Statistics

To calculate a correlation between user's perceived realism and the *Haptic Fidelity* of a haptic feedback system we first needed to calculate the effect size of perceived realism between conditions in the 38 selected papers. For the effect size we chose the common metric of the correlation coefficient (Pearson's r), as it is a versatile effect size metric and widely used, e.g., by Cummings and Bailenson [13]. The correlation coefficient was mainly derived from the reported means and standard deviations of the measured perceived realism in each condition from the papers. In cases where this data was not reported, the correlation coefficient was derived from the reported t, F and χ^2 statistics with only one degree of freedom [56]. An effect size was calculated between all conditions of each paper, indicating the effect that the compared haptic feedback methods had on the perceived realism. To calculate the correlation with the *Haptic Fidelity* we used the difference in *Haptic Fidelity* scores for the two haptic feedback methods from the compared conditions and the respective effect size between these conditions. We calculated the final correlation coefficient (Pearson's r) between perceived realism and the *Haptic Fidelity* based on all calculated effect sizes and corresponding differences in *Haptic Fidelity* scores.

5.4 Results

The calculated correlation coefficient between perceived realism and the *Haptic Fidelity* is $r(155) = 0.69, p < .00001$ with a 95 % confidence interval from 0.6 to 0.76 (see Figure 6) and an aggregated sample size of $K = 703$. This indicates that the *Haptic Fidelity* score of the framework is strongly positively correlated to user's self-reported perceived realism of the haptic feedback method according to Cohen [11, 12].

A substantial number of papers (17 of 38) use a standard VR controller with vibration as one of the feedback conditions. One could argue that it is obvious that the effect sizes as well as the differences in *Haptic Fidelity* scores differ greatly between standard controllers with vibrations and elaborate haptic feedback systems and therefore attribute greatly to the strong correlation. To evaluate if the strong correlation is influenced by this comparison, we did a second analysis excluding all conditions that used a standard VR controller with vibration for haptic feedback. For this analysis, we excluded papers that only had two conditions where one was a standard controller (9) and for all other papers we only compared conditions that used other feedback methods than standard controllers. In this analysis 29 papers with 128 conditions were included. The correlation was calculated the same way as described before. For this second analysis we found a correlation coefficient between perceived realism and the *Haptic Fidelity* of $r(110) = 0.8, p < .00001$ with a 95 % confidence interval from 0.72 to 0.86 and an aggregated sample size of $K = 536$. This shows an even stronger correlation

[2]The complete dataset of scores for each of the 154 conditions of the selected research papers is provided in the supplementary materials.

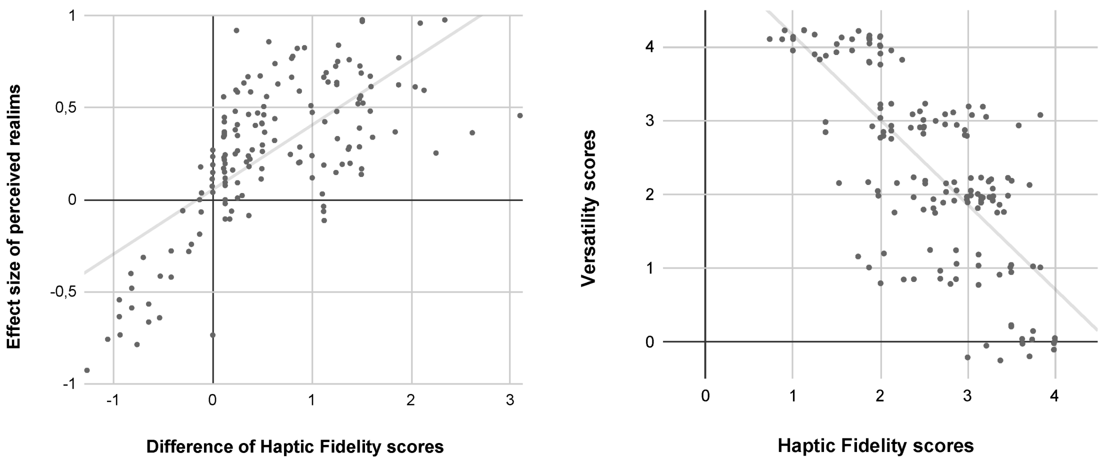


Figure 6: Left: scatter plot with effect size of perceived realism between 154 conditions from the 38 assessed papers and the corresponding difference between the *Haptic Fidelity* scores of the same conditions. Right: scatter plot of the *Haptic Fidelity* and *Versatility* scores from the 38 assessed papers (154 conditions) with a random jitter of .25 for better visualization of the discrete categories for the *Versatility.*

between the *Haptic Fidelity* score and user's self-reported perceived realism of the haptic feedback method than before.

Finally, to investigate the relationship between the *Haptic Fidelity* and *Versatility* dimension of the framework we also calculated the correlation between the *Haptic Fidelity* and the corresponding *Versatility* scores for all feedback systems. We found a correlation coefficient of $r(152) = -0.72, p < .00001$ with a 95 % confidence interval from -0.79 to -0.64 (see Figure 6), indicating a strong negative correlation between the *Haptic Fidelity* and *Versatility* of haptic feedback systems.

We provide all data of the analysis, including the extracted data of all papers, calculated effect sizes, *Haptic Fidelity* and *Versatility* scores and final correlation coefficients in the supplementary material.

6 DISCUSSION

The results of the analysis indicate that the *Haptic Fidelity* dimension of the framework is an indicator for the perceived realism of the haptic feedback in VR confirming our first hypothesis. This makes the *Haptic Fidelity* dimension a valid measure of immersion describing the potential of a system to enable an elevated user experience. Based on the additional analysis, in which we omitted the conditions with vibration feedback of standard VR controllers, we can conclude that the *Haptic Fidelity Framework* is suited for all kinds of haptic feedback systems, especially as there were remarkably diverse feedback systems included in the analysis. The strong negative correlation between *Haptic Fidelity* and *Versatility* scores shows that there is a trade-off between the two dimensions confirming our second hypothesis. A similar trade-off between the realism and versatility of interactions was also identified by Jacob et al. [29] suggesting that this is a common trade-off encountered in interactive systems.

When looking at current systems (see Figure 5), there are no systems falling into the category of abstract but specific feedback. First, this is because abstract feedback is naturally not specific as it can be repurposed to represent a variety of other more specific haptic impressions and second, there is no practical value for designers to build such systems. On the opposite side of the spectrum there are also no systems yet that fall into the category of very realistic but also very generic feedback. Such systems which create realistic feedback for any use case present the ultimate goal of haptic feedback for VR and could someday lead to systems like the Holodeck from Star Trek[3]. The design space offered by our framework should help researchers and designers to understand and compare current feedback systems but also give direction for the development of future haptic devices.

Haptic feedback occurs as a reaction to the interaction with a system and therefore the two are intricately connected. In our framework we try to separate the two as strictly as possible by only looking at output related metrics to solely capture factors influencing the haptic feedback. The interaction can be analyzed in a similar fashion with the framework for interaction fidelity analysis (FIFA) by McMahan et al. [51]. The interplay between haptic feedback and interaction stands out when integrating results from both frameworks (see Figure 2). The FIFA framework for example examines the "kinetic symmetry" describing the involvement of the same forces as in the real-world action. These forces could be created by a haptic feedback system and are described by the *Stimuli* and *Magnitude* factors of our framework. To design and build VR systems that are perceived as extremely realistic both, the interaction and haptic feedback, and their interplay must be considered. Beyond the relation to the user experience and interaction, there are many more factors that are important measures for haptic

[3]https://intl.startrek.com/database_article/holodeck

feedback systems, e.g., cost, portability, setup time, accessibility that can be considered.

With *Haptic Fidelity* our framework provides a measure of immersion assessing perceptual and technical aspects of haptic feedback for VR. While the validation has shown that it is a predictor for one aspect of user experience (perceived realism), it does not cover how users actually experience the feedback. The actual user experience of haptics is based on many factors, e.g., state of mind and past experiences, which can be quite individual for each user [42]. In addition, the haptic perception does not have the same spatial and temporal precision as other senses like vision, making it even more prone to individual interpretation. These aspects of user's individual haptic experience are not captured by this framework. Instead, it is meant to provide researchers and designers with the knowledge which factors can potentially have an influence on the user experience. When rating the individual factors of the framework it is also important to separate the experience part from the actual objective criteria, which should be evaluated. For our validation, we argue that rating the systems by experts, who did not design nor try the systems themselves[4], provides an assessment that is focused on the actual perceptual and technological qualities and not on the subjective experience of using the system. This, of course, requires a comprehensive description about the intent of a system (i.e., the application scenario), the haptic feedback provided and the underlying hardware and software. In our analysis, this information was taken from prototype descriptions in publications and accompanying videos. Within our analysis, the experts had to intensively engage with the information about the systems and in some cases, the experts' judgments could vary. We acknowledge that assessments made with this framework are of a qualitative nature and that the scores provide the means for structuring the results. While the underlying perceptual and technological measures are objective, their assessments naturally are subjective and can be subject to discussion. The ratings for the validation were made by experienced researchers who agreed on a consistent basis of assessment in their discussions forming a solid foundation for our validation process. We also affirm that classifying haptic feedback for a specific application on the single scales requires a certain level of experience with haptic systems, the range of conceivable feedback that could be provided, as well as how the scales integrate this. To address this, we provide detailed explanations with example classification and justifications for each scale of the framework as well as a full documentation of the 154 example conditions that were rated in the supplementary materials for reference. Nevertheless, even with potential small differences in the resulting haptic fidelity score by different evaluators, the framework is valuable for unfolding the influence of the diverse parameters that contribute to haptic fidelity and provides means for comparing different approaches in a structured way.

With this paper we contribute to the general understanding of haptic feedback. While we specifically target VR systems and focus on measures of immersion that are closely related to the field of VR, the definition of *Haptic Fidelity* and the individual factors we identified could as well be applied in general to all kinds of haptic feedback systems. However, one characteristic of haptics in headset-based VR is that haptic and visual feedback can be provided independently, in contrast to haptic feedback in real environments, where the visual appearance of haptic devices or passive haptic props is usually directly perceived by the user. In line with VR-based systems, our framework focuses on aspects of realism of isolated haptic sensing, independently of the visual appearance of the used artifact.

Even though one focus of this paper is to find a single comprehensive measure for *Haptic Fidelity* to compare systems and validate the framework, we emphasize the importance of the individual factors that are presented in this paper. They provide the means to form a deep and structured understanding of underlying concepts of haptic feedback and inform decisions when researching and designing haptic feedback systems. Thus, the *Haptic Fidelity Framework* offers potential to support researchers, designers and practitioners in a variety of situations such as making informed design decisions for haptic feedback and exploring the haptic "design space", estimating potential differences in perceived realism when comparing haptic devices (which in parts could be already possible on a concept level before the devices are actually built), finding detailed explanations for these differences, improving existing haptic devices or addressing the trade-off between realistic and generic feedback.

We acknowledge the following limitations for this work: The types of systems that were analyzed for the validation were not equally distributed, e.g., there were more conditions with custom-built controllers and little to no papers with full-body feedback systems like exoskeletons. This might be due to the existence of less papers on these type of systems and the selection criteria we applied. In addition, when the researchers read the papers to rate the *Haptic Fidelity* of systems for our validation, they could have read or at least looked at the results section of the papers. Even though we encouraged them to not look at results, we did not black them out and it would have been possible to read them. Looking at the results or even graphs about the perceived realism could have influenced the ratings and introduced a bias in the analysis. Furthermore, there might be a general publication bias towards papers with significant differences between haptic feedback conditions, which could have led to an increased number of papers with large effects for our validation.

7 CONCLUSION

In this paper we present the *Haptic Fidelity Framework* providing the means for a structured, comprehensive, and in-depth understanding of factors that influence the realism of haptic feedback in virtual reality. It allows to assess all types of haptic feedback systems for VR. The framework describes the level of sensory fidelity and the resulting capability to produce realistic or abstract feedback in the *Haptic Fidelity* dimension, containing 14 fine-grained factors. A second *Versatility* dimension represents how specific the haptic feedback of a system is for the particular application or if it is more generic and can potentially cover a wider range of applications. We validate our framework by applying it to 154 haptic feedback conditions of 38 research papers on virtual reality applications and compare the *Haptic Fidelity* score to the reported perceived realism.

[4]From the 38 prototypes that were assessed in the analysis, some of the experts had personally experienced four systems before.

The results show a strong correlation suggesting that the framework (1) is well suited to assess haptic feedback systems in VR and (2) describes the potential of a system to create realistic feedback for users. Additionally, we found a strong negative correlation between *Haptic Fidelity* and *Versatility* indicating that most current feedback systems make the trade-off between highly realistic but application-specific and more abstract but widely applicable feedback. While this framework is based on the assessment of perceptual and technological immersion effects, the subjective user experience is also of immense importance when analyzing haptic feedback systems. As current measures, e.g., presence questionnaires, cover haptic feedback only to a minimal degree, in future work we aim to develop a dedicated questionnaire assessing the user experience of haptic feedback systems.

ACKNOWLEDGMENTS

This research is partially funded by the German Federal Ministry of Education and Research (BMBF) in the grant program "Gesundes Leben" (healthy living) and the Klaus Tschira Stiftung.

REFERENCES

[1] Imtiaj Ahmed, Ville Harjunen, Giulio Jacucci, Eve Hoggan, Niklas Ravaja, and Michiel M. Spapé. 2016. Reach out and Touch Me: Effects of Four Distinct Haptic Technologies on Affective Touch in Virtual Reality. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction* (Tokyo, Japan) *(ICMI '16)*. Association for Computing Machinery, New York, NY, USA, 341–348. https://doi.org/10.1145/2993148.2993171

[2] Amy L Alexander, Tad Brunyé, Jason Sidman, and Shawn A Weil. 2005. From gaming to training: A review of studies on fidelity, immersion, presence, and buy-in and their effects on transfer in pc-based simulations and games. *DARWARS Training Impact Group* 5 (2005), 1–14.

[3] Mahdi Azmandian, Mark Hancock, Hrvoje Benko, Eyal Ofek, and Andrew D. Wilson. 2016. *Haptic Retargeting: Dynamic Repurposing of Passive Haptics for Enhanced Virtual Reality Experiences*. Association for Computing Machinery, New York, NY, USA, 1968–1979. https://doi.org/10.1145/2858036.2858226

[4] Christopher C. Berger, Mar Gonzalez-Franco, Eyal Ofek, and Ken Hinckley. 2018. The uncanny valley of haptics. *Science Robotics* 3, 17 (2018), eaar7010. https://doi.org/10.1126/scirobotics.aar7010 arXiv:https://robotics.sciencemag.org/content/3/17/eaar7010.full.pdf

[5] Mette Boldt, Michael Bonfert, Inga Lehne, Melina Cahnbley, Kim Korschinq, Loannis Bikas, Stefan Finke, Martin Hanci, Valentin Kraft, Boxuan Liu, Tram Nguyen, Alina Panova, Ramneek Singh, Alexander Steenbergen, Rainer Malaka, and Jan Smeddinck. 2018. You Shall Not Pass: Non-Intrusive Feedback for Virtual Walls in VR Environments with Room-Scale Mapping. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. Institute of Electrical and Electronics Engineers, New York, NY, USA, 143–150. https://doi.org/10.1109/VR.2018.8446177

[6] Barbara Boos and Henning Brau. 2017. Erweiterung des UEQ um die Dimensionen Akustik und Haptik. In *Mensch und Computer 2017 - Usability Professionals*, Steffen Hess and Holger Fischer (Eds.). Gesellschaft für Informatik e.V., Regensburg, 7 pages. https://doi.org/10.18420/muc2017-up-0236

[7] Felix Born, Maic Masuch, and Antonia Hahn. 2020. Ghost Sweeper: Using a Heavy Passive Haptic Controller to Enhance a Room-Scale VR Exergame. In *2020 IEEE Conference on Games (CoG)*. Institute of Electrical and Electronics Engineers, New York, NY, USA, 221–228. https://doi.org/10.1109/CoG47356.2020.9231867

[8] Elodie Bouzbib, Gilles Bailly, Sinan Haliyo, and Pascal Frey. 2021. "Can I Touch This?": Survey of Virtual Reality Interactions via Haptic Solutions. *CoRR* abs/2101.11278 (2021), 16 pages. arXiv:2101.11278 https://arxiv.org/abs/2101.11278

[9] Doug A. Bowman and Ryan P. McMahan. 2007. Virtual Reality: How Much Immersion Is Enough? *Computer* 40, 7 (2007), 36–43. https://doi.org/10.1109/MC.2007.257

[10] Shaoyu Cai, Pingchuan Ke, Takuji Narumi, and Kening Zhu. 2020. ThermAirGlove: A Pneumatic Glove for Thermal Perception and Material Identification in Virtual Reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. Institute of Electrical and Electronics Engineers, New York, NY, USA, 248–257. https://doi.org/10.1109/VR46266.2020.00044

[11] Jacob Cohen. 1992. Statistical power analysis. *Current directions in psychological science* 1, 3 (1992), 98–101.

[12] Jacob Cohen. 2013. *Statistical power analysis for the behavioral sciences*. Academic press, New York, NY, USA.

[13] James J Cummings and Jeremy N Bailenson. 2016. How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media Psychology* 19, 2 (2016), 272–309.

[14] Donald Degraen, André Zenner, and Antonio Krüger. 2019. *Enhancing Texture Perception in Virtual Reality Using 3D-Printed Hair Structures*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300479

[15] Thomas Fröhlich, Dmitry Alexandrovsky, Timo Stabbert, Tanja Döring, and Rainer Malaka. 2018. VRBox: A Virtual Reality Augmented Sandbox for Immersive Playfulness, Creativity and Exploration. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play* (Melbourne, VIC, Australia) *(CHI PLAY '18)*. Association for Computing Machinery, New York, NY, USA, 153–162. https://doi.org/10.1145/3242671.3242697

[16] Dominik Gall and Marc Erich Latoschik. 2018. The Effect of Haptic Prediction Accuracy on Presence. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. Institute of Electrical and Electronics Engineers, New York, NY, USA, 73–80. https://doi.org/10.1109/VR.2018.8446153

[17] Guilherme Gonçalves, Miguel Melo, José Vasconcelos-Raposo, and Maximino Bessa. 2020. Impact of Different Sensory Stimuli on Presence in Credible Virtual Environments. *IEEE Transactions on Visualization and Computer Graphics* 26, 11 (2020), 3231–3240. https://doi.org/10.1109/TVCG.2019.2926978

[18] Martin Grunwald. 2008. *Human haptic perception: Basics and applications*. Springer Science & Business Media, Basel, Switzerland.

[19] Sebastian Günther, Mohit Makhija, Florian Müller, Dominik Schön, Max Mühlhäuser, and Markus Funk. 2019. PneumAct: Pneumatic Kinesthetic Actuation of Body Joints in Virtual Reality Environments. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (San Diego, CA, USA) *(DIS '19)*. Association for Computing Machinery, New York, NY, USA, 227–240. https://doi.org/10.1145/3322276.3322302

[20] Tobias Günther, Lars Engeln, Sally J. Busch, and Rainer Groh. 2019. The Effect of Elastic Feedback on the Perceived User Experience and Presence of Travel Methods in Immersive Environments. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. Institute of Electrical and Electronics Engineers, New York, NY, USA, 613–620. https://doi.org/10.1109/VR.2019.8798119

[21] Abdelwahab Hamam, Mohamad Eid, Abdulmotaleb El Saddik, and Nicolas D. Georganas. 2008. A Quality of Experience Model for Haptic User Interfaces. In *Proceedings of the 2008 Ambi-Sys Workshop on Haptic User Interfaces in Ambient Media Systems* (Quebec City, Canada) *(HAS '08)*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), Brussels, BEL, Article 1, 6 pages.

[22] Ping-Hsuan Han, Yang-Sheng Chen, Kong-Chang Lee, Hao-Cheng Wang, Chiao-En Hsieh, Jui-Chun Hsiao, Chien-Hsing Chou, and Yi-Ping Hung. 2018. Haptic around: Multiple Tactile Sensations for Immersive Environment and Interaction in Virtual Reality. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology* (Tokyo, Japan) *(VRST '18)*. Association for Computing Machinery, New York, NY, USA, Article 35, 10 pages. https://doi.org/10.1145/3281505.3281507

[23] Vincent Hayward and Oliver R. Astley. 1996. Performance Measures for Haptic Interfaces. In *Robotics Research*, Georges Giralt and Gerhard Hirzinger (Eds.). Springer London, London, 195–206.

[24] David Hecht and Miriam Reiner. 2009. Sensory dominance in combinations of audio, visual and haptic stimuli. *Experimental brain research* 193, 2 (2009), 307–314.

[25] Ronan Hinchet, Velko Vechev, Herbert Shea, and Otmar Hilliges. 2018. DextrES: Wearable Haptic Feedback for Grasping in VR via a Thin Form-Factor Electrostatic Brake. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology* (Berlin, Germany) *(UIST '18)*. Association for Computing Machinery, New York, NY, USA, 901–912. https://doi.org/10.1145/3242587.3242657

[26] Matthias Hoppe, Pascal Knierim, Thomas Kosch, Markus Funk, Lauren Futami, Stefan Schneegass, Niels Henze, Albrecht Schmidt, and Tonja Machulla. 2018. VRHapticDrones: Providing Haptics in Virtual Reality through Quadcopters. In *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia* (Cairo, Egypt) *(MUM 2018)*. Association for Computing Machinery, New York, NY, USA, 7–18. https://doi.org/10.1145/3282894.3282898

[27] Inwook Hwang, Hyungki Son, and Jin Ryong Kim. 2017. AirPiano: Enhancing music playing experience in virtual reality with mid-air haptic feedback. In *2017 IEEE World Haptics Conference (WHC)*. Institute of Electrical and Electronics Engineers, New York, NY, USA, 213–218. https://doi.org/10.1109/WHC.2017.7989903

[28] Brent Edward Insko. 2001. *Passive haptics significantly enhances virtual environments*. The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.

[29] Robert J.K. Jacob, Audrey Girouard, Leanne M. Hirshfield, Michael S. Horn, Orit Shaer, Erin Treacy Solovey, and Jamie Zigelbaum. 2008. Reality-Based Interaction: A Framework for Post-WIMP Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. Association for Computing Machinery, New York, NY, USA, 201–210. https://doi.org/10.1145/1357054.1357089

[30] Caroline Jay, Mashhuda Glencross, and Roger Hubbold. 2007. Modeling the Effects of Delayed Haptic and Visual Feedback in a Collaborative Virtual Environment. *ACM Trans. Comput.-Hum. Interact.* 14, 2 (Aug. 2007), 8–es. https://doi.org/10.1145/1275511.1275514

[31] Seungwoo Je, Myung Jin Kim, Woojin Lee, Byungjoo Lee, Xing-Dong Yang, Pedro Lopes, and Andrea Bianchi. 2019. Aero-Plane: A Handheld Force-Feedback Device That Renders Weight Motion Illusion on a Virtual 2D Plane. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 763–775. https://doi.org/10.1145/3332165.3347926

[32] Lynette A Jones. 2000. Kinesthetic sensing. In *in Human and Machine Haptics*. Citeseer, MIT Press, Cambridge, MA, USA, 10 pages.

[33] Lynette A. Jones and Hong Z. Tan. 2013. Application of Psychophysical Techniques to Haptic Research. *IEEE Transactions on Haptics* 6, 3 (2013), 268–284. https://doi.org/10.1109/TOH.2012.74

[34] Namkyoo Kang and Sangwon Lee. 2018. A Meta-Analysis of Recent Studies on Haptic Feedback Enhancement in Immersive-Augmented Reality. In *Proceedings of the 4th International Conference on Virtual Reality* (Hong Kong, Hong Kong) *(ICVR 2018)*. Association for Computing Machinery, New York, NY, USA, 3–9. https://doi.org/10.1145/3198910.3198911

[35] Astrid M.L. Kappers and Wouter M. Bergmann Tiest. 2013. Haptic perception. *WIREs Cognitive Science* 4, 4 (2013), 357–374. https://doi.org/10.1002/wcs.1238 arXiv:https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/wcs.1238

[36] Thorsten A Kern. 2009. *Engineering haptic devices: a beginner's guide for engineers*. Springer Publishing Company, Incorporated, New York, NY, USA.

[37] Mohamed Khamis, Nora Schuster, Ceenu George, and Max Pfeiffer. 2019. ElectroCutscenes: Realistic Haptic Feedback in Cutscenes of Virtual Reality Games Using Electric Muscle Stimulation. In *25th ACM Symposium on Virtual Reality Software and Technology* (Parramatta, NSW, Australia) *(VRST '19)*. Association for Computing Machinery, New York, NY, USA, Article 13, 10 pages. https://doi.org/10.1145/3359996.3364250

[38] Erin Kim and Oliver Schneider. 2020. *Defining Haptic Experience: Foundations for Understanding, Communicating, and Evaluating HX*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376280

[39] Robert Kovacs, Eyal Ofek, Mar Gonzalez Franco, Alexa Fay Siu, Sebastian Marwecki, Christian Holz, and Mike Sinclair. 2020. *Haptic PIVOT: On-Demand Handhelds in VR*. Association for Computing Machinery, New York, NY, USA, 1046–1059. https://doi.org/10.1145/3379337.3415854

[40] Andrey Krekhov, Katharina Emmerich, Philipp Bergmann, Sebastian Cmentowski, and Jens Krüger. 2017. *Self-Transforming Controllers for Virtual Reality First Person Shooters*. Association for Computing Machinery, New York, NY, USA, 517–529. https://doi.org/10.1145/3116595.3116615

[41] Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and Evaluation of a User Experience Questionnaire. In *HCI and Usability for Education and Work*, Andreas Holzinger (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 63–76.

[42] Effie Lai-Chong Law, Virpi Roto, Marc Hassenzahl, Arnold P.O.S. Vermeeren, and Joke Kort. 2009. Understanding, Scoping and Defining User Experience: A Survey Approach. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 719–728. https://doi.org/10.1145/1518701.1518813

[43] Susan J Lederman and Roberta L Klatzky. 2009. Haptic perception: A tutorial. *Attention, Perception, & Psychophysics* 71, 7 (2009), 1439–1459.

[44] Robert W. Lindeman and Steffi Beckhaus. 2009. Crafting Memorable VR Experiences Using Experiential Fidelity. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology* (Kyoto, Japan) *(VRST '09)*. Association for Computing Machinery, New York, NY, USA, 187–190. https://doi.org/10.1145/1643928.1643970

[45] Mark Lindquist, Bruce Maxim, Jennifer Proctor, and Francine Dolins. 2020. The effect of audio fidelity and virtual reality on the perception of virtual greenspace. *Landscape and Urban Planning* 202 (2020), 103884. https://doi.org/10.1016/j.landurbplan.2020.103884

[46] Pedro Lopes, Alexandra Ion, and Patrick Baudisch. 2015. Impacto: Simulating Physical Impact by Combining Tactile Stimulation with Electrical Muscle Stimulation. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (Charlotte, NC, USA) *(UIST '15)*. Association for Computing Machinery, New York, NY, USA, 11–19. https://doi.org/10.1145/2807442.2807443

[47] Pedro Lopes, Sijing You, Lung-Pan Cheng, Sebastian Marwecki, and Patrick Baudisch. 2017. *Providing Haptics to Walls & Heavy Objects in Virtual Reality by Means of Electrical Muscle Stimulation*. Association for Computing Machinery, New York, NY, USA, 1471–1482. https://doi.org/10.1145/3025453.3025600

[48] Emanuela Maggioni, Erika Agostinelli, and Marianna Obrist. 2017. Measuring the added value of haptic feedback. In *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*. Institute of Electrical and Electronics Engineers, New York, NY, USA, 1–6. https://doi.org/10.1109/QoMEX.2017.7965670

[49] Ryan Patrick McMahan. 2011. *Exploring the effects of higher-fidelity display and interaction for virtual reality games*. Ph.D. Dissertation. Virginia Tech.

[50] Ryan P. McMahan, Doug A. Bowman, David J. Zielinski, and Rachael B. Brady. 2012. Evaluating Display Fidelity and Interaction Fidelity in a Virtual Reality Game. *IEEE Transactions on Visualization and Computer Graphics* 18, 4 (2012), 626–633. https://doi.org/10.1109/TVCG.2012.43

[51] Ryan P McMahan, Chengyuan Lai, and Swaroop K Pal. 2016. Interaction fidelity: the uncanny valley of virtual reality interactions. In *International Conference on Virtual, Augmented and Mixed Reality*. Springer, Springer International Publishing, Cham, 59–70.

[52] James L Meriam, L Glenn Kraige, and Jeff N Bolton. 2020. *Engineering mechanics: dynamics*. John Wiley & Sons, Hoboken, NJ, USA.

[53] Thomas Muender, Anke V. Reinschluessel, Sean Drewes, Dirk Wenig, Tanja Döring, and Rainer Malaka. 2019. *Does It Feel Real? Using Tangibles with Different Fidelities to Build and Explore Scenes in Virtual Reality*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300903

[54] Marianna Obrist, Sue Ann Seah, and Sriram Subramanian. 2013. *Talking about Tactile Experiences*. Association for Computing Machinery, New York, NY, USA, 1659–1668. https://doi.org/10.1145/2470654.2466220

[55] Andrea Passalenti, Razvan Paisa, Niels C. Nilsson, Nikolaj S. Andersson, Federico Fontana, Rolf Nordahl, and Stefania Serafin. 2019. No Strings Attached: Force and Vibrotactile Feedback in a Virtual Guitar Simulation. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. Institute of Electrical and Electronics Engineers, New York, NY, USA, 1116–1117. https://doi.org/10.1109/VR.2019.8798168

[56] Robert Rosenthal and M Robin DiMatteo. 2001. Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual review of psychology* 52, 1 (2001), 59–82.

[57] Andreas Ryge, Lui Thomsen, Theis Berthelsen, Jonatan S. Hvass, Lars Koreska, Casper Vollmers, Niels C. Nilsson, Rolf Nordahl, and Stefania Serafin. 2017. Effect on high versus low fidelity haptic feedback in a virtual reality baseball simulation. In *2017 IEEE Virtual Reality (VR)*. Institute of Electrical and Electronics Engineers, New York, NY, USA, 365–366. https://doi.org/10.1109/VR.2017.7892328

[58] Mikel Sagardia and Thomas Hulin. 2018. Multimodal Evaluation of the Differences between Real and Virtual Assemblies. *IEEE Transactions on Haptics* 11, 1 (2018), 107–118. https://doi.org/10.1109/TOH.2017.2741488

[59] Evren Samur. 2012. *Performance Evaluation Based on Physical Measurements*. Springer London, London, 43–65. https://doi.org/10.1007/978-1-4471-4225-6_4

[60] Antti Sand, Ismo Rakkolainen, Poika Isokoski, Jari Kangas, Roope Raisamo, and Karri Palovuori. 2015. Head-Mounted Display with Mid-Air Tactile Feedback. In *Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology* (Beijing, China) *(VRST '15)*. Association for Computing Machinery, New York, NY, USA, 51–58. https://doi.org/10.1145/2821592.2821593

[61] Suji Sathiyamurthy, Melody Lui, Erin Kim, and Oliver Schneider. 2021. Measuring Haptic Experience: Elaborating the HX model with scale development. In *2021 IEEE World Haptics Conference (WHC)*. Institute of Electrical and Electronics Engineers, New York, NY, USA, 979–984. https://doi.org/10.1109/WHC49131.2021.9517220

[62] Oliver Schneider, Karon MacLean, Colin Swindells, and Kellogg Booth. 2017. Haptic experience design: What hapticians do and where they need help. *International Journal of Human-Computer Studies* 107 (2017), 5–21.

[63] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. 2001. The experience of presence: Factor analytic insights. *Presence: Teleoperators & Virtual Environments* 10, 3 (2001), 266–281.

[64] Peter Schulz, Dmitry Alexandrovsky, Felix Putze, Rainer Malaka, and Johannes Schöning. 2019. *The Role of Physical Props in VR Climbing Environments*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300413

[65] Hasti Seifi, Farimah Fazlollahi, Michael Oppermann, John Andrew Sastrillo, Jessica Ip, Ashutosh Agrawal, Gunhyuk Park, Katherine J. Kuchenbecker, and Karon E. MacLean. 2019. *Haptipedia: Accelerating Haptic Device Discovery to Support Interaction & Engineering Design*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300788

[66] Hasti Seifi, Michael Oppermann, Julia Bullard, Karon E. MacLean, and Katherine J. Kuchenbecker. 2020. Capturing Experts' Mental Models to Organize a Collection of Haptic Devices: Affordances Outweigh Attributes. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376395

[67] Daniel Shor, Bryan Zaaijer, Laura Ahsmann, Simon Immerzeel, Max Weetzel, Daniël Eikelenboom, Jess Hartcher-O'Brien, and Doris Aschenbrenner. 2018. Designing Haptics: Comparing Two Virtual Reality Gloves with Respect to Realism, Performance and Comfort. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. Institute of Electrical and Electronics Engineers, New York, NY, USA, 318–323. https://doi.org/10.1109/ISMAR-Adjunct.2018.00095

[68] Mel Slater. 2003. A Note on Presence Terminology. *Presence Connect* 3 (01 2003).

[69] Mel Slater and Anthony Steed. 2000. A virtual presence counter. *Presence* 9, 5 (2000), 413–434.

[70] Anthony Steed, Sebastian Friston, Vijay Pawar, and David Swapp. 2020. Docking Haptics: Extending the Reach of Haptics by Dynamic Combinations of Grounded and Worn Devices. In *26th ACM Symposium on Virtual Reality Software and Technology* (Virtual Event, Canada) *(VRST '20)*. Association for Computing Machinery, New York, NY, USA, Article 2, 11 pages. https://doi.org/10.1145/3385956.3418943

[71] Evan Strasnick, Christian Holz, Eyal Ofek, Mike Sinclair, and Hrvoje Benko. 2018. *Haptic Links: Bimanual Haptics for Virtual Reality Using Variable Stiffness Actuation.* Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3174218

[72] Yuqian Sun, Shigeo Yoshida, Takuji Narumi, and Michitaka Hirose. 2019. *PaCaPa: A Handheld VR Device for Rendering Size, Shape, and Stiffness of Virtual Objects in Tool-Based Interactions.* Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300682

[73] Yudai Tanaka, Arata Horie, and Xiang 'Anthony' Chen. 2020. DualVib: Simulating Haptic Sensation of Dynamic Mass by Combining Pseudo-Force and Texture Feedback. In *26th ACM Symposium on Virtual Reality Software and Technology* (Virtual Event, Canada) *(VRST '20)*. Association for Computing Machinery, New York, NY, USA, Article 1, 10 pages. https://doi.org/10.1145/3385956.3418964

[74] Shan-Yuan Teng, Tzu-Sheng Kuo, Chi Wang, Chi-huan Chiang, Da-Yuan Huang, Liwei Chan, and Bing-Yu Chen. 2018. PuPoP: Pop-up Prop on Palm for Virtual Reality. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology* (Berlin, Germany) *(UIST '18)*. Association for Computing Machinery, New York, NY, USA, 5–17. https://doi.org/10.1145/3242587.3242628

[75] Hsin-Ruey Tsai, Ching-Wen Hung, Tzu-Chun Wu, and Bing-Yu Chen. 2020. *ElastOscillation: 3D Multilevel Force Feedback for Damped Oscillation on VR Controllers.* Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376408

[76] Hsin-Ruey Tsai, Jun Rekimoto, and Bing-Yu Chen. 2019. *ElasticVR: Providing Multilevel Continuously-Changing Resistive Force and Instant Impact Using Elasticity for VR.* Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3290605.3300450

[77] Martin Usoh, Ernest Catena, Sima Arman, and Mel Slater. 2000. Using presence questionnaires in reality. *Presence* 9, 5 (2000), 497–503.

[78] Yuntao Wang, Zichao (Tyson) Chen, Hanchuan Li, Zhengyi Cao, Huiyi Luo, Tengxiang Zhang, Ke Ou, John Raiti, Chun Yu, Shwetak Patel, and Yuanchun Shi. 2020. *MoveVR: Enabling Multiform Force Feedback in Virtual Reality Using Household Cleaning Robot.* Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376286

[79] Eric Whitmire, Hrvoje Benko, Christian Holz, Eyal Ofek, and Mike Sinclair. 2018. *Haptic Revolver: Touch, Shear, Texture, and Shape Rendering on a Reconfigurable Virtual Reality Controller.* Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173660

[80] Bob G Witmer and Michael J Singer. 1998. Measuring presence in virtual environments: A presence questionnaire. *Presence* 7, 3 (1998), 225–240.

[81] Jackie (Junrui) Yang, Hiroshi Horii, Alexander Thayer, and Rafael Ballagas. 2018. VR Grabbers: Ungrounded Haptic Retargeting for Precision Grabbing Tools. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology* (Berlin, Germany) *(UIST '18)*. Association for Computing Machinery, New York, NY, USA, 889–899. https://doi.org/10.1145/3242587.3242643

[82] Yuan-Syun Ye, Hsin-Yu Chen, and Liwei Chan. 2019. Pull-Ups: Enhancing Suspension Activities in Virtual Reality with Body-Scale Kinesthetic Force Feedback. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 791–801. https://doi.org/10.1145/3332165.3347874

[83] Yan Yixian, Kazuki Takashima, Anthony Tang, Takayuki Tanno, Kazuyuki Fujita, and Yoshifumi Kitamura. 2020. *ZoomWalls: Dynamic Walls That Simulate Haptic Infrastructure for Room-Scale VR World.* Association for Computing Machinery, New York, NY, USA, 223–235. https://doi.org/10.1145/3379337.3415859

[84] Shigeo Yoshida, Yuqian Sun, and Hideaki Kuzuoka. 2020. *PoCoPo: Handheld Pin-Based Shape Display for Haptic Rendering in Virtual Reality.* Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376358

[85] André Zenner and Antonio Krüger. 2019. *Drag:On: A Virtual Reality Controller Providing Haptic Feedback Based on Drag and Weight Shift.* Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300441

[86] André Zenner and Antonio Krüger. 2017. Shifty: A Weight-Shifting Dynamic Passive Haptic Proxy to Enhance Object Perception in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* 23, 4 (2017), 1285–1294. https://doi.org/10.1109/TVCG.2017.2656978

[87] Kening Zhu, Taizhou Chen, Feng Han, and Yi-Shiun Wu. 2019. *HapTwist: Creating Interactive Haptic Proxies in Virtual Reality Using Low-Cost Twistable Artefacts.* Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300923

Publication P3

Examining Design Choices of Questionnaires in VR User Studies

Dmitry Alexandrovsky, Susanne Putze, Michael Bonfert, Sebastian Höffner, Pitt Michelmann, Dirk Wenig, Rainer Malaka, and Jan David Smeddinck

This publication investigates embedding questionnaires in virtual environments (inVRQs) as a self-reporting method for user studies. We discuss a literature review revealing 15 recent studies using inVRQs and present a survey with 67 VR experts from academia and industry. Based on the outcomes, we conducted two user studies to test the usability of different presentation and interaction methods of inVRQsand compare them to conventional questionnaires outside VR (outVRQs). We observed comparable completion times with higher enjoyment but lower usability of inVRQs. Our findings advocate the application of inVRQs and inform the design considerations.

My contribution: I contributed to the systematic literature review (*investigation*), to the research idea and prototype design (*conceptualization*) and study design (*methodology*) of the main user study, as well as to the writing and revision of the manuscript (*writing – draft, review & editing*).

Published in the *Proceedings of the 2020 ACM Conference on Human Factors in Computing Systems (CHI '20)*

Examining Design Choices of Questionnaires in VR User Studies

Dmitry Alexandrovsky*,[1], Susanne Putze*,[1], Michael Bonfert[1], Sebastian Höffner[1], Pitt Michelmann[1], Dirk Wenig[1], Rainer Malaka[1], Jan David Smeddinck[2]
[1] Digital Media Lab, University of Bremen, Germany
[2] Open Lab, School of Computing, Newcastle University, UK
{dimi,sputze,bonfert,shoeffner,pitt,dwenig,malaka}@uni-bremen.de Jan.Smeddinck@newcastle.ac.uk

ABSTRACT
Questionnaires are among the most common research tools in virtual reality (VR) user studies. Transitioning from virtuality to reality for giving self-reports on VR experiences can lead to systematic biases. VR allows to embed questionnaires into the virtual environment which may ease participation and avoid biases. To provide a cohesive picture of methods and design choices for questionnaires in VR (INVRQ), we discuss 15 INVRQ studies from the literature and present a survey with 67 VR experts from academia and industry. Based on the outcomes, we conducted two user studies in which we tested different presentation and interaction methods of INVRQS and evaluated the usability and practicality of our design. We observed comparable completion times between INVRQS and questionnaires outside VR (OUTVRQS) with higher enjoyment but lower usability for INVRQS. These findings advocate the application of INVRQS and provide an overview of methods and considerations that lay the groundwork for INVRQ design.

Author Keywords
Virtual reality; VR; user studies; in-VR questionnaires; inVRQs; research methods.

CCS Concepts
•**Human-centered computing → Virtual reality; HCI design and evaluation methods;** *Empirical studies in HCI;* User studies;

INTRODUCTION
The notable rise of a new generation of virtual reality (VR) systems in recent years opened up new methods and interventions for researchers across many different areas. These range from highly immersive stimulus-response studies [36, 60] over spatial navigation [147, 175] and embodied cognition [149, 154] to exposure therapies [47, 56, 146], exercising [165, 202], education [18, 104], work collaboration [7, 97], and other forms of social interaction [5, 99]. Typically, mid- and post-experience measures are collected via subjective responses on questionnaires [103]. Furthermore, the development and evaluation of VR experiences for entertainment or serious purposes also frequently relies on questionnaires. To fill out conventional paper- or computer-based forms, the subjects need to take off the head-mounted display (HMD) and change the domain from virtual to physical reality [83]. This often leads to temporal disorientation and loss of sense of control [91]. Accordingly, questionnaire results are likely biased to a degree that is difficult to quantify and likely varies from case to case. Such undetermined bias is highly problematic for many types of research and evaluations.

In contrast to the physical domain, alternate reality technologies allow for the embedding of questionnaires directly into the environment. While the transition from VR to answering, for example, paper-based questions presents a drastic change of context, embedding question-items in VR offers an opportunity to stay closer to the context of an ongoing exposure than out-of-VR research setups and avoid a break in presence (BIP) [83, 137]. Especially for measures where the self-reporting needs to be administered as soon as possible after the treatment and may be especially sensitive to differences in study setups, such as common measures for presence [155], immersion [82] or flow [35, 183], it appears crucial to give careful consideration to the interaction modalities around delivering question-items and gathering responses in order to foster the adequate interpretation of individual research outcomes and for fostering replicability.

Schwind et al. observed a higher consistency of self-reported presence when administering questionnaires in VR. The authors highlight that the effects of using questionnaires in VR are unclear, pointing out that the commonly applied measures were not validated for VR studies [159]. These considerations motivate our investigation on questionnaire practices in contemporary VR user research. We investigate whether authors employ comparable terminology and reflect their choices with regard to questionnaire presentation and response collection mechanisms. To date, VR user research does not have a shared range of common administration procedures, well-defined classification schemes, or standardized toolkits for presenting questionnaires in VR user studies that could guide

*The authors contributed equally.

CHI '20, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6708-0/20/04 ...$15.00.
http://dx.doi.org/10.1145/3313831.3376260

such considerations and make it easier to contrast against other work.

To build an understanding of the current practices and to lay the groundwork for future efforts around developing practical toolkits, we based this work on the following research questions:

RQ1 *How are questionnaires and individual question-items applied in contemporary VR user research?*
RQ2 *What are the considerations around – and possible advantages of – administering questionnaires in VR?*
RQ3 *What are appropriate design choices for presentation and interaction with in-VR questionnaires?*

To approach the research questions, we conducted: (i) a literature review of 123 publications on VR user studies, (ii) a survey with 74 VR experts, (iii) a preliminary design study to compare different presentation and interaction methods, (iv) a user study (n=38) of a revised iteration of our in-VR questionnaire tool, comparing it to a on-screen response system. This aggregated examination of the literature review and expert survey allows us to extract a detailed picture of proceedings of VR user research, compensating for incomplete reporting in publications and for sampling effects [123]. Based on these insights, we iteratively implemented an in-VR questionnaire tool and applied it in 2 user studies to investigate design options and effects of questionnaire administration in VR.

This work focuses on self-reporting methods. Other more objective measures (e.g. behavioural or biosignals) provide more reliable data [79] and are less likely to be affected by BIPs. We also did not take qualitative analyses into account since the procedures differ notably from quantitative studies. By providing a coherent survey of questionnaire methods in VR through a literature analysis, expert interviews, and an empirical investigation of in-VR questionnaire (INVRQ) designs, our research can inform the discourse around questionnaire methods in VR research, structure design considerations on VR user research and also inform the design and implementation of practical questionnaire toolkits that are relevant to both research and industry.

STATE OF THE ART

Due to its immersive nature and a wide variety in technical setups, VR with HMDs requires careful deliberation by researchers aiming to conduct studies with human subjects. In this section, we review methods and practices for questionnaires in human-subject research followed by a consideration of the VR-specific technicalities around moving between worlds and their effects on question asking.

Questionnaires in Human Subject Research

Questionnaires are an important source of information for evidence-based research [12, 45, 103]. They embody self-reports and therefore gather the participants' subjective experiences [45]. Question types in surveys can be divided into unstructured and structured questions [144]. Structured questions allow for a clear classification of the responses (see [152] for a more detailed discussion), while unstructured or open-ended questions allow the subjects to respond freely. This type of question is more exploratory and allows for a broader understanding of phenomena [103], while also requiring more effort from the respondents. Survey methodologies received much attention in the literature and their advantages or drawbacks are widely explored [13, 17, 163]. *Reliable* (consistency of the measurement) and *validated* (measuring the right construct) questionnaires are vital for reproducible and consistent research [17]. Choi and Pak [32] list 3 groups of potential biases: *design of the question*, *questionnaire design*, and *administration*. Question design covers the effects of poor wording, such as double-barreled questions, negative phrasing or wording that enforces choices [12, 32, 103]. Biases of questionnaire design are due to formatting and length of the surveys [4, 32] as well as length and structure of the questions [23]. Context-dependent forgetting [1, 58] due to environment change [136] biases the responses. We argue that especially in immersive scenarios, a series of random errors can be minimized through consistent administration of questionnaires. Notably, these considerations on questionnaire design are typically contextualized against paper-based or screen-based questionnaires not considering aspects around BIPs or switching between different realities [159].

Moving Between Virtual and Physical Reality

When individuals are deeply engaged with an activity or absorbed in a virtual environment (VE), they completely block out the world around them [35]. Brown and Cairns [26] identified 3 levels of immersion in games as a scale of involvement: (i) *engagement*, the lowest level of immersion, (ii) *engrossment*, when players become emotionally affected by the VR and (iii) *total immersion* where players are detached from reality. This phase is also associated with empathy for the characters in the game and transfer of consciousness [150] and is linked to the state of flow [35]. A sudden interruption or transition between realities can invoke negative feelings and affect the emotional state [91]. Accordingly, assessing presence during immersive experiences results in more reliable measurements [21, 49]. In contrast to immersion, presence is a state of mind, describing the feeling of being part of the VE [82, 201]. Presence relies on involvement and immersion [201]. When "returning" from a state of presence in VEs, a real world task is impaired to the degree of immersion and one's ability to re-engage with the "real world" is decreased [82, 171].

Thus, leaving the VE is likely to interrupt the presence perception. Schwind et al. [159] investigated the effect of filling out a questionnaire directly in VR. They replicated their lab in VR and asked participants to fill out presence questionnaires in physical reality and in VR after exposing participants to a VE at varying degrees of realism. Schwind et al. could not find significant differences of presence between the 2 questionnaire modalities. However, the data revealed a lower variance and, thus, higher consistency of the data when the questions were answered in VR. This is in line with evidence from the literature that support the assessment of questions in VR [83, 91].

Frommel et al. observed a significant effect on presence when questionnaires are integrated in the game context [53]. Similarly, Shute discussed how to embed questionnaires into games without disturbing the game flow [169]. These considerations

With Q, not reported: [2, 15, 16, 20, 27, 30, 31, 33, 34, 38, 42, 43, 50–52, 54, 63, 67–70, 73–76, 86, 89, 90, 92, 93, 95, 101, 105–107, 109–111, 115, 117, 120, 122, 124–126, 130–132, 134, 135, 139–143, 145, 151, 156, 164, 167, 173, 174, 178–181, 184–186, 188, 190, 192, 193, 195, 199, 200, 204]
OUTVRQS: [3, 8, 29, 77, 84, 94, 98, 114, 121, 138, 157, 162, 176, 177, 187, 191, 196, 197]
INVRQS: [6, 28, 37, 44, 55, 66, 85, 100, 108, 127, 148, 158–160, 198]
Without Q: [9, 11, 19, 22, 40, 59, 71, 91, 133, 153, 182, 194, 205]

Table 1. Overview of all papers reviewed, organized by whether a questionnaire was used and how it was presented

can guide the choice of adequate levels of *embeddedness*, and a general decision whether to remain in VR or to exit VR when asking study participants to respond to questionnaires or individual question-items.

LITERATURE REVIEW

For our literature review we searched in digital libraries [10, 80] that host proceedings of high quality conference series about human-centric methods and VR applications. Namely, we scanned ACM CHI, ACM CHI PLAY, ACM VRST, IEEE VR and IEEE 3DUI for the years 2016–2019. We deliberately chose this timeframe to gain consistent insights on contemporary resarch methods and the release of the HTC Vive in June 2016 [78] substantially changed HMD use in VR user studies. The search included publications until July 2019, considering papers categorized with any of the keywords "virtual reality", "head mounted display", "virtual environment", "user study", and "questionnaire". Only papers with abstracts that mention a VR application with HMDs and explicitly mention some form of evaluation with users or empirical user study were added to the list.

In total we reviewed and evaluated 123 research papers, see Table 1. The papers are organized following whether they use questionnaires for measurement or not. We also categorized the different representations of questionnaires: in-VR questionnaires (INVRQ), out-VR questionnaires (OUTVRQ) and those who did not include a report of the presentation method.

Analysis

To investigate common procedures of questionnaires in VR user research, we analyzed both the system design and the study design. We paid special attention to the treatment of questionnaires or individual question-items and to the description regarding transitions between VR and physical reality. Further, we looked at the nature of the VR experience. These factors were considered as discriminatory for VR user research from classic, less immersive interventions. Each paper was examined by 2 of the authors. Disagreements were resolved in discussions.

Uses of VR Hardware

Established desktop VR devices, such as the HTC Vive (63) and Oculus Rift (42), are most commonly used in the papers reviewed followed by mobile device powered HMDs, e.g. Samsung Gear VR (9) and Google Cardboard (4). 3 publications did not report the device used. The input modalities and devices used in the VR applications are mainly native VR controllers (e.g. Oculus Touch, HTC Vive Controller) (49) followed by freehand interaction (e.g. Leap Motion or Microsoft Kinect) (35) and general purpose input devices (e.g. game controller, keyboard, mouse, stylus, smartwatch, and touch screen) (25).

	Interaction	Presentation	Questionnaire (Extent)
[6]	Gamepad	n.a.	well-being [44] (SI)
[28]	Gamepad	HUD (Fig. 1b)	well-being [44] (SI)
[37]	Orally	HUD	custom (MI)
[44]	Gamepad	World (Fig. 1e)	well-being (SI)
[55]	n.a	n.a.	presence rating [21] (SI)
[66]	VR controller	World	well-being [44] (SI)
[85]	VR controller	HUD (Fig. 1a)	custom (SI)
[100]	VR controller	World	custom (SI)
[108]	Full body	n.a.	IAT [64] (MI)
[127]	VR controller	World (Fig. 1d)	NASA-TLX [72] (MI)
[148]	Freehand	n.a.	custom (MI)
[158]	Freehand	World (Fig. 1c)	PQ [201] (MI)
[159]	VR controller	World (Fig. 1f)	SUS [189], IPQ [155], PQ [201] (MI)
[160]	Freehand	World (Fig. 1c)	PQ [201] (MI)
[198]	Freehand	Body (Fig. 1g)	PANAS [96] (MI)

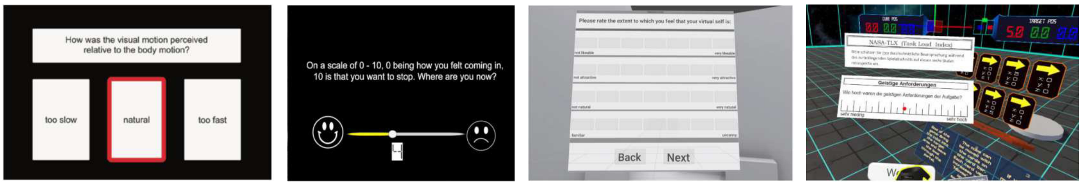
Table 2. Examples of INVRQS with their realization (interaction and presentation), the questionnaire used and its extent (multi-items (MI) vs. single-item (SI)), if reported in the publication (n.a. otherwise)

Questionnaire Assessment

110 out of 123 papers report having used questionnaires in their VR user studies. Since the use of VR devices entails design decisions regarding the presentation of questionnaires and individual question-items, we surveyed the documentation of such decisions in the respective papers. 77/110 do not report how they presented the questionnaires to their users. 13 papers report that the participants filled out the questionnaire after leaving the VE but do not describe whether they used paper- or screen-based questionnaires. 15 papers report on the usage of INVRQS – either for the whole question asking procedures in the user-study (3) or in combination with OUTVRQS (12).

Cases of In-VR Questionnaires

15 papers report the use of INVRQS. Some describe the design in more detail. Figure 1 depicts 7 different realizations of INVRQS. Kang et al. (Fig. 1a) used a 2D heads-up display (HUD) overlay with a single question about the user's motion perception between multiple trials in their VE [85]. The user interface (UI) shows a single question with a multi-line question text and 3 buttons for answering choices. As input device, they used a native VR controller [85]. Schwind et al. (Fig. 1c) included the full 32-item PQ [201]. The participants stayed in the VE for the whole duration of the study: on average 58.6 min [160] and 75 min [158]. The authors designed a 3D floating UI which appears in front of the subjects showing a one-line text instruction and 4 items on 7-point Likert-scales. Users select answers and navigate the questionnaire with freehand gestures using a Leap Motion [158, 160, 161]. In another study, Schwind et al. (Fig. 1f) placed single questions on presence on a virtual PC in the VE, with which a user interacts with the trackpad of a VR controller [159]. Oberdörfer et al. (Fig. 1d) presented the NASA-TLX [72] using a virtual world-referenced representation of the paper-based version. The users interacted using a VR controller with pointing [127].



(a) Kang et al. [85] (b) Cao et al. [28] (c) Schwind et al. [161] (d) Oberdörfer et al. [127]

(e) Fernandes and Feiner [44] (f) Schwind et al. [159] (g) Wienrich et al. [198]

Figure 1. Examples of different realizations of INVRQ: (a) and (b) present the questionnaire using a HUD, (c)-(f) use a world-referenced questionnaire, and (g) presents the questionnaire attached to the body.

Wienrich et al. (Fig. 1g) presented a body-referenced INVRQ [198]. The questionnaire was displayed on a 2D floating UI with a 20-items PANAS [96] attached to the hand of the virtual character. They combined the in-experience measurement with further out-VR measurements and oral answering of the Fast Motion Sickness Scale (FMS) [88]. Fernandes and Feiner (Fig. 1e) is the earliest example of INVRQS in our sample. The authors applied a 10-point Likert-scale slider on well-being where subjects could stop the experiment by selecting the maximum value of discomfort [44]. [28] (Fig. 1b), [6] and [66] adapted this method in their works in different realizations.

All INVRQ designs differ in their presentation (HUD, world-reference, body-referenced), their extent (single-question vs. multi-item questionnaire), question-item presentation (text-based vs. scales) and interaction modality (pointing, free-hand, trackpad). Table 2 summarizes the variation of the INVRQ designs. The applications of INVRQS cover questions about the subject's well-being, their sense of presence and task-specific questionnaires, e.g. about task workload or affect.

Discussion

Our literature analysis shows that comprehensive reporting of questionnaire usage in VR research is frequently neglected. Only Schwind et al. [159] compare INVRQS and OUTVRQS and discuss their effects. This indicates that the field may benefit from building awareness and providing guidelines. We identified 15/123 cases of INVRQS applied in VR user studies. The realizations differ substantially in their presentations and interaction methods. The majority of cases used presentations that contextualized the questionnaires in the VE, either attached to the user, or anchored in world space in a stationary manner. The participants used predominantly native controllers to interact in VR, directly followed by free-hand interaction. In order to better contextualize these literature findings and to collect details about the design of INVRQS and their potential shortcomings, we discuss a further investigation through an online expert survey in the following section.

EXPERT SURVEY

To augment the insights gained from the literature review, we conducted an expert online survey that evaluates general proceedings of VR user studies and attitudes towards INVRQS. With this additional analysis of the state of the art, we aimed to capture an impression of the actual procedures employed by the researchers independent from possible biases present in publications with space limitations.

Survey Dissemination and Pre-Processing

We developed a custom survey to capture the general reporting of proceedings of VR user studies as well as experiences and attitudes of the study designers regarding INVRQS. The complete survey is provided in an OSF project[1]. Following informed-consent, it consisted of 22 questions grouped into 5 categories (demographics, general research practice, VR research practice, INVRQ experience and OUTVRQ experience). The survey was designed and distributed using Google Forms. For recruiting expert participants, we extracted a list of authors from the papers (2016–2018) we analyzed in our literature review and sent them personal invitations. We also advertised the survey via social media channels. Over 6 weeks in July and August 2018, we collected 74 replies.

Since the online survey focused on researchers and VR experts, we excluded 4 participants who were not directly involved in VR user studies in the last 24 months and 3 participants who indicated that they do not generally use questionnaires as measures in their VR user studies. Moreover, we corrected obvious spelling mistakes to facilitate accurate counts of established terms (e.g. hardware or questionnaire names). For anonymized analysis, we removed time stamps and added unique identifiers (E1–E74). The reported analysis focuses only on questions that are most relevant to our research questions.

Analysis

After data cleaning, the analyzed data set consisted of 67 full set responses. Based on participant indication, we sampled

[1] https://osf.io/f5qy7/

responses from 13 different *countries of residence* (Q3). With 20 participants from Germany, 9 responses each from the USA and Portugal and 4 responses each from Denmark and France, the majority of our sample came from northern-hemisphere countries. As a result of the sampling strategy, the expert sample stems predominantly from academia. The participants indicated to be holding the following positions (Q2): 22 Ph.D. students, 16 professors (full, associate, assistant or equivalent), 10 undergraduate students, and 8 Postdocs. The remaining 11 participants provided individual answers, such as research engineer or research fellow. On average, the participants indicated to be rather experienced with designing user studies (Q4). On a 6-point Likert-scale (0 to 5): M=3.76, SD=1.18. The majority (56) were directly involved with conducting 1−5 VR user studies in the last 24 month while 8 experts indicated involvement with between 6 and 10 studies. Only 3 indicated more than 10 VR user studies (Q9). Our sample consists of a diverse spectrum within the group of academic researchers who have a sound background on VR user research allowing us to interpret the responses as an expert evaluation.

VR research

The experts' *most commonly used VR devices* (Q10, multiple choice) in the last 24 months were *desktop VR devices* (HTC Vive (55) and Oculus Rift (37)) or *mobile VR devices* (Samsung Gear VR (15) and Google Cardboard (11)). Other devices were mentioned occasionally. Only 4 participants used neither HTC Vive nor Oculus Rift. The most commonly used input modalities (Q11, multiple choice) reported for the general interaction with VR are mainly *native VR controllers* (50) and *freehand interaction* (e.g. Leap motion or Microsoft Kinect) (31) followed by *gaze interaction* (19) and *general purpose input devices* (e.g. game controller, keyboard, mouse, stylus or smartwatch) (34). 6 participants reported using custom controllers that were not further defined.

Questionnaires

We asked our participants how they *usually present questionnaires* in their user studies. 29 use a separate screen outside the VR, 21 paper questionnaires, 6 embed the questionnaire in VR, 2 use oral answers and 9 use mixed methods (Q13). Validated questionnaires (OUTVRQS) are very common in VR user research (Q12). Independent of the form of presentation, the experts reported they *encounter some difficulties with questionnaires in VR user studies* (Q16): 14 experts reported problems with the questionnaires in general, such as ambiguous question items in validated questionnaires, e.g. PQ, or the length of questionnaires, e.g. SSQ [87], NASA-TLX [72], especially for use in measurements between trials. 7 experts reported problems of their participants when moving from VR to reality for answering the questionnaires, namely the lack of immersion (4) and the temporal effort (3).

29/67 experts have tried in-VR questionnaires (INVRQ *users*). Most of them (25) tried fully interactive questionnaires that are both presented and answered in VR (Q18, single choice). All experts rated the usefulness of INVRQS (Q17) on a 6-point Likert scale (0−5). They rated the usefulness with M=2.97, SD=1.58 which is significantly higher than the midpoint 2.5 (one-sided independent sample t-test t_{65}=2.41,

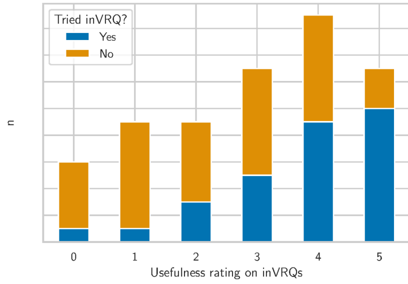


Figure 2. Ratings of the usefulness of INVRQS (Q17, scale: 0 to 5, 5 being highest) separated by INVRQ experience (Yes/No).

p<.01). For further analysis, the experts were split into 2 sub-groups according to whether they have tried an INVRQ in their research setups (Q18) or not (M_{Yes}=3.72, SD_{Yes}=1.33 and M_{No}=2.39, SD_{No}=1.55), cf. Figure 2. A two-sided independent sample t-test showed a significant difference between prior INVRQ experience and the usefulness rating (t_{65}=3.69, p<.01, Cohen's d=.45). To appropriately weight the responses we further split the experts according to their scientific seniority. Group A: post-docs and professors, Group B: all other participants (presumably more junior researchers). The experts in group A tend to rate the usefulness higher than the other participants: With M_A=3.46, SD_A=1.74 and M_B=2.70, SD_B=1.46). But the difference was not significant t_{65}=1.91, p=.06.

In-VR Questionnaires

14/29 INVRQ *users* presented the questions in world space as floating displays and 4 used a 2D UI (probably replacing the virtual world) (Q19). 4 experts responded that they presented only 1 question at a time. 14 experts did not observe any difficulties of their participants using INVRQS (Q20). Usability and the "amount of time for filling out the questionnaires" (E44) were major challenges experts came across (4× each). Further, "completing long questionnaires can be (...) annoying" (E18) and participants are "impatient (...) and hence made mistakes in the scaled questions" (E5).

Out-VR Questionnaires

Using Q18 as a branching question, 38 OUTVRQ *users* were asked to provide reasons for not embedding the questionnaires into the VE (Q21). 19 from that group reported that they do not suspect to benefit from INVRQS. 9 experts reported technical difficulties and 3 exceeding time effort in setting up the embedding for INVRQS. "Time effort to embed into VR much higher than doing the same on paper and harder to change/make additions" (E48). 6 experts mentioned usability problems and 1 sees problems in all these categories. The experts explained their critical opinions of INVRQS further (Q22): They argue, it would reduce the "willingness to answer the questions" (E31) because interaction with the UI in VR is more frustrating and takes longer than on desktop. Especially text entry is seen as a major issue. The experts reason that this might bias the results "towards the choices that are easiest to make through the interface" (E31). Additionally, they mentioned the time effort for implementation and usage:

"My last study took 45 minutes including questionnaires. It gets very hot and inconvenient under a Vive for such a long time – especially if it's your first time" (E71).

Discussion

The responses to the survey indicate that experts generally appear to have a positive attitude towards INVRQS. Both our literature review and expert survey demonstrate that there is no common standard for using questionnaires in VR user studies. While the reporting of the presentation type of the questionnaires in the literature is mostly imprecise, our expert survey did not surface a clear standard method for presentation. Thus, regarding RQ1, the state of the art for delivering questionnaires in VR user studies is quite heterogeneous. Also, we found inconsistencies between the literature analysis and the online survey: While only 15/123 papers reported on the usage of questionnaires embedded in VR, 29/67 experts reported that they have tried INVRQS before. Possible reasons for the contrast between the positive attitude of the INVRQ user group and the rare reporting of INVRQS are not intuitively clear. The difference could indicate that a shift may be taking place in the community with many applications of INVRQS not having been published yet. There is, however, the possibility of increased variance due to publication bias that warrants control in future work. Alternatively, the authors of the analyzed corpus may have considered the presentation type to be irrelevant. In contrast to these outcomes, there are experts who have a strong opinion against embedding questionnaires and argue that INVRQS could bias responses. In response to RQ2, the broad range of *reasons provided for or against using* INVRQS indicate technical challenges, implementation effort, and fear of biases and participant overload. In the following section we investigate these objections from the users' viewpoint and design and evaluate an INVRQ tool.

DESIGN STUDY

Although many guidelines from "traditional" HCI, cf. [41, 168], can be applied to VR, the third dimension brings its own challenges for UI designers. LaViola et al. [102] point out that there are no standards for 3D UIs yet, such as the ISO 9241-110 [81] standards for 2D GUIs. The results from the literature review and from the expert surveys are heterogeneous and do not allow for a definite answer of the usefulness and realization of INVRQS. With regard to RQ3, the goal of this design study was to identify the most usable interface design.

In-VR Questionnaire Design

Based on the results of our expert survey and general guidelines on traditional UIs [41, 168] as well as UIs in VR [40,62,128,203], we implemented 4 variants of INVRQS. The interfaces differ in 2 dimensions: *anchoring* (world-referenced and body-referenced) and *interaction modality* (pointer and trackpad). The world-referenced UI is anchored in the VE and users can freely move around the interface. A body-referenced UI is attached to a hand-held controller.

To enhance readability, we applied guidelines from research and industry which recommend *signed distance field* fonts [61, 128]. In line with Dingler et al. [40], we used light glyphs on a dark background. Oculus best practices recommend to avoid HUDs (cf. Fig. 1a, 1b) in favor of UI elements that are settled in the VE, as it overcomes the binocular disparity and allows to contextualize the UI in VR [102, 128]. This is in line with the statements from the experts who applied INVRQS, as the majority (32/67) implemented questionnaires in the world space (see Fig. 1c–1f). In contrast to world-referenced UIs, body-referenced interfaces, as in Fig. 1g, take advantage of the user's proprioceptive sense and can significantly enhance the interaction with the UI [102, 118].

The interaction with the UI varies between a laser pointer – as Oberdörfer et al. [127] applied for their INVRQS – and a clicking interaction where the trackpad of the controller is used to navigate through the questionnaires, similar to Schwind et al. [159]). Oculus guidelines [128] suggest a laser pointer with a visible ray-cast and a cursor projected on the UI as an appropriate and intuitive method to interact with UI in VR. This is reflected in general guidelines which suggest better performance in terms of speed, accuracy and cognitive demand [112, 113, 119]. In contrast, navigating a UI using a trackpad promises to be more efficient when the UI is close to the users since they are not required to twist their arms for aiming.

The 4 designs that emerge from the 2 dimensions anchoring (world, body) and interaction (pointer, trackpad) cover a wide range of designs that are applicable to INVRQS. We developed all 4 designs iteratively following the same usability guidelines and paying particularly attention for comparability. Our interface supports continuous values (slider), checklists, radio lists, drop downs; and switches. The prototype is implemented in Unity3D with OpenVR. Their interaction and design is demonstrated in the accompanying video figure.

Study Design, Procedure and Tasks

The study has a 2×2 within-subject design where users filled out INVRQS using the 4 different versions of the interface: world-pointer (WP), world-trackpad (WT), body-pointer (BP) and body-trackpad (BT). The order of the condition was counterbalanced using Latin Square. First, the subjects were welcomed and informed about the study. The experimenter fit the HMD (HTC Vive) and explained the interaction in the current condition with the native controllers. To provide a context to the VE, we used a sci-fi scene.

To investigate the usability of all response types, we developed a questionnaire which asks for common knowledge facts to ask subjects easy-to-answer but objective questions to calculate correctness. Each condition consisted of a questionnaire that comprised all 5 question types once. Each participant answered 20 questions in total with each question type once per condition. After each condition in VR, the participants took off the HMD and filled out a paper-based System Usability Scale (SUS) [25]. When the participants finished all 4 conditions, the experimenter asked them to put the 4 interfaces into a ranked order and conducted an interview.

Participants

10 male (age M=29.9, SD=2.9) subjects from a game jam at the campus participated in the study. For their experience in VR and as VR developer or researcher, the participants gave

a score on a Likert scale from 1 (no experience) to 5 (high experience). All participants had prior experience in VR and game development (*M*=3.6, *SD*=.84). 6 participants developed VR applications or conducted studies in VR (*M*=2.3, *SD*=1.33). Although this sample lacks representative diversity, it allows for a deep discussion of the design space while avoiding general issues that could have come up with novices.

Results and Discussion

For the quantitative metrics (SUS [25] and duration), we conducted a repeated measures ANOVA with condition as factor. The descriptive statistics and the results of the analysis are in Table 3. Bonferroni corrected pairwise comparisons of the SUS [25] scores revealed significant differences between BP and BT ($p<.01$), between BP and WP ($p<.01$), and between WP and WT ($p<.01$). For the required time, Bonferroni corrected pairwise comparisons revealed significant differences between BP and BT ($p<.01$), BT and WP ($p<.01$), and between WP and WT ($p<.05$). On the subjective rankings, the participants rated the world-pointer setup significantly higher ($F_{3,36}=13.27$, $p<.001$).

6/10 subjects stated that the laser pointer is easy and intuitive to use. In contrast, the majority perceived the trackpad as tedious and confusing. 3/10 participants liked being able to move the questionnaire. But 4/10 participants reported they were confused by the body-referenced interface and stated the movement as unpleasant. The world-anchored UI was stated as less demanding than the body-referenced (3/10).

Completion time, SUS scores [25] and the interviews clearly show the world-referenced anchoring with the laser pointer interaction is easiest to use and therefore the best candidate to investigate the concerns raised by the experts.

USER STUDY

To assess whether usability and duration concerns about INVRQS raised by the experts hold true and to provide guidance on question-asking methods in VR, we conducted a user study in which the participants shot balloons with bow and arrow in a VR archery game and then filled out INVRQS as well as questionnaires presented on a notebook (OUTVRQ). The aim of this study was to evaluate adequate design choices of INVRQS we identified in the design study and to capture the users' perspectives on INVRQS. In contrast to previous work by Schwind et al. [159], our user study applies the questionnaires in a realistic study setting without replicating the lab space virtually.

The Questionnaire Tools

We implemented the INVRQ with world-based anchoring and laser pointer interaction (Fig. 3). We refined the overall readability, i.e. font size, contrast, spacing and positioning

	WP (M, SD)	WT (M, SD)	BP (M, SD)	BT (M, SD)	F(3,27)
SUS [25]	91.25, 8.99	64.25, 24.09	79.00, 14.49	52.25, 19.52	11.15**
t (min)	0.61, 0.12	1.00, 0.27	0.69, 0.33	1.19, 0.23	14.92**

Table 3. Descriptive statistics and RM-ANOVA of the design study for the 4 conditions world-pointer (WP), world-trackpad (WT), body-pointer (BP) and body-trackpad (BT) on SUS [25] and time. **p<.01.

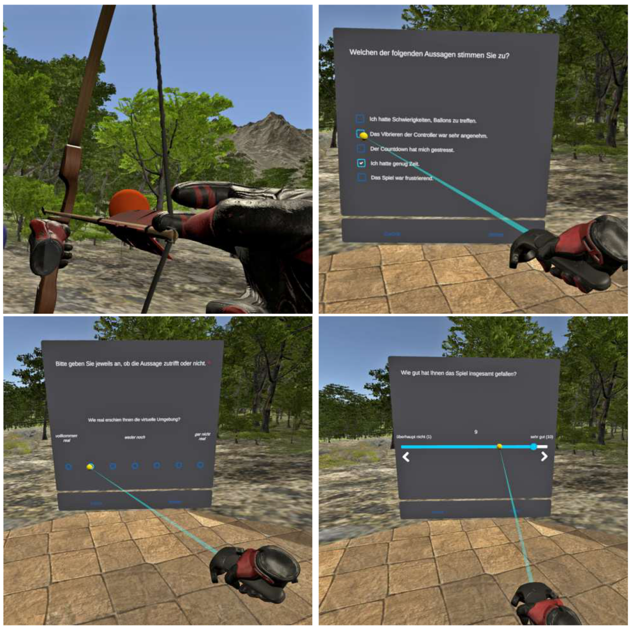


Figure 3. Screenshots of the archery task. The world space-anchored INVRQ is filled out using the HTC Vive controller as a laser pointer.

of the UI in the VE. The tool supports sliders, radio lists, radio grids and check lists. We omitted drop downs and free-text fields since these elements rarely appear in standardized questionnaires, as the literature review shows. If needed, drop downs can be represented using a radio list, and free-text input can be approximated with voice recording.

The OUTVRQS were realized using the questionnaire tool LimeSurvey [57]. It was presented on a 15" notebook with external keyboard and mouse with default speed.

Measurements

After the archery game, we measured presence in the VE using IPQ [155] on a Likert scale with the subscales *general presence (GP)*, *spatial presence (SP)*, *involvement (INV)* and *experienced realism (REAL)*. Furthermore, we asked the participants to rate the game and the perceived control over the bow on a 10-ticks slider. For demonstrating a greater variety of question types, we additionally included questions about the VE (1*x* numerical, 4*x* single choice with 2–5 items, 4*x* multiple choice with 5 and 16 items).

To evaluate the workload and the usability of the INVRQ tool, we employed the raw NASA-TLX [72], using a 20-ticks slider to measure physical, cognitive and temporal demand, as well as performance, effort, and frustration. Usability was measured with the UMUX [46], a four-item questionnaire providing comparability with the SUS [25] (r=.96). As an objective performance metric, we logged the exact time. To get detailed insights from the users' perspective, we conducted a semi-structured interview at the end of a session.

Study Design and Procedure

The study compares the usage of INVRQS to digital OUTVRQS. To provide a realistic study setting, we designed a

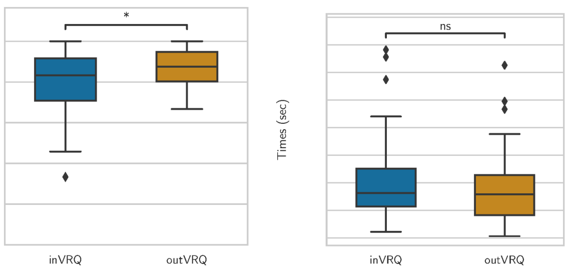


Figure 4. Ratings of usability on UMUX [155] (left) and of completion times (right) for both conditions.

balloon archery game as an immersive VR experience (see Fig. 3). We chose this task because (i) it is engaging and requires the participants to focus, promising an immersive VR experience, (ii) the interaction is easy to learn but differs from the pointing interaction used for answering the INVRQS; thus, the task is less likely to produce carry-over learning effects. For the archery task, we used free Unity3D assets and implementations from the SteamVR Interaction System. The VE consists of a round platform with 3 pillars that display instructions that guide through the experiment. The platform is surrounded by 12 spawn points for the balloon targets and it is situated in a realistic environment with mountains, trees, a river and high resolution textures. The questionnaire interaction builds on the laser pointer by SteamVR and the GUI interaction by HTC Vive.

The study followed a within-subject design with the conditions INVRQ and OUTVRQ in randomized order. After the participants were informed about the procedure, they signed a consent form. The experimenter fitted the HTC Vive Pro HMD. Then the participants played a tutorial round in which they had to hit 5 balloons followed by a 90 s round where they should hit as many balloons as possible. After the archery task, the participants filled out questionnaires using the corresponding tool for their first condition (INVRQ, OUTVRQ). Afterwards, the participants repeated the game and questionnaires using the other questionnaire method. We encouraged them to take a 2 min break in between. After both conditions, the experimenter conducted a semi-structured interview followed by a paper-based demographics questionnaire. Finally, the participants were orally debriefed. The study, including game, questionnaires and interview, was conducted in German and took around 45 min ($\approx$11 min in VR).

Participants

We advertised the study on campus, social media and in lectures and conducted it in July 2019. In total, 38 participants (age: M=27, SD=10.8; 16 f, 22 m, 0 other; 20 started with the INVRQ condition) volunteered for our study. Most participants were students. 21 participants used vision aids in VR, 1 participant has a light dyschromatopsia. The sample has a broad range of prior VR experience: 6 participants use VR regularly, 27 occasionally and 5 never used VR before. 19/38 participated in other VR user studies previously and 3 participants used INVRQS before. We detected no outliers regarding demographics and task completion times.

Results

Performance, Presence and Rating of the VR Experience

On average, participants shot 24.71 balloons (SD=10.33) in the first round and improved by 6.34 (SD=6.01) balloons in the second round. We obtained presence on the IPQ [155] on a 7-point Likert scale (0–6). To determine if the measures deviate from neutral, we performed a two-sided one-sample t-tests against the midpoint 3. The results show a positive difference for GP (t_{37}=14.93, p<.001), SP (t_{37}=22.53, p<.001) and INV (t_{37}=5.99, p<.001), but no difference for REAL ($t_{37}=-48$, p=.63). There was no significant effect of condition on any IPQ subscales (see Fig. 5a). Like Schwind et al. [159], we performed a t-test to compare the variances of both conditions on all IPQ subscales; the differences were not significant (GP_{var}: t_{37}=-.52, p=.60, SP_{var}: t_{37}=-.05, p=.96, INV_{var}: t_{37}=1.47, p=.15 $REAL_{var}$: t_{37}=.17, p=.87).

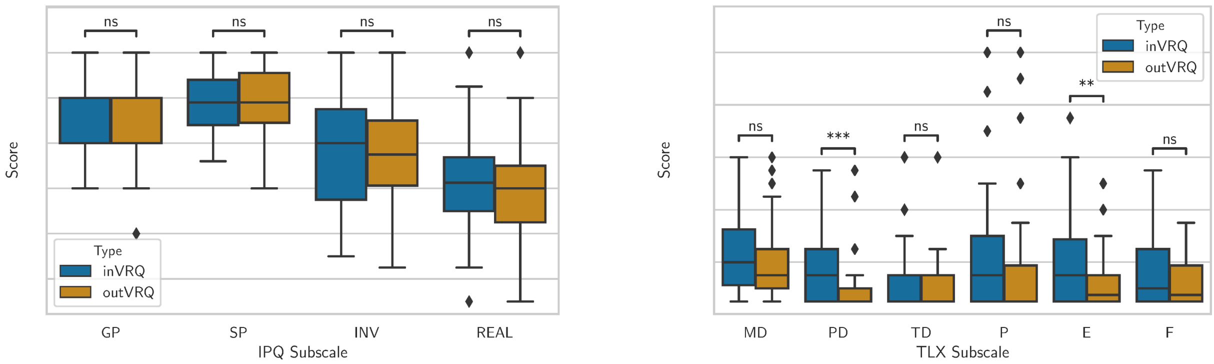
Further, the participants rated the interaction with the bow (M=7.89, SD=1.70) and how they liked the game (M=8.0, SD=1.73) on a 10-ticks analog scale from 1 (not at all) to 10 (very much). A one-sample t-test revealed significant differences against the midpoint (5.5) for bow control (t_{37}=15.01, p<.001) and game ratings (t_{37}=14.86, p<.001), but no effect of conditions neither for bow control nor game rating.

Duration and Self-reports on Usage of the Questionnaires

Figun 4 shows plots of UMUX scores and answering time. On average, participants required 6.77 min (SD=2.69) to fill out the questionnaires in VR and 6.30 min (SD=2.26) on the notebook (without the time for taking off the HMD). There was no significant difference of condition on duration of filling out the questionnaires (t_{37}=-1.05, p=.29). Nevertheless, in the post-experiment interviews, 5 participants perceived filling out the questionnaires in VR as faster. 2 had the impression the questionnaires in VR would be shorter. The IPQ completion times did not differ significantly (INVRQ IPQ: M=146.39 s, SD=65.46) in comparison to the ones reported by Schwind et al. (VR IPQ: M=146.94 s, SD=63.2) [159]: t_{37}=.04, p=.97. On the UMUX, the participants rated the questionnaire tools' usability. Both systems were rated positively (INVRQ: M=77.35, SD=18.35, OUTVRQ: M=86.21, SD=9.4). The difference between the conditions was significant (t_{37}=2.82, p<.01). In accordance with Grier [65], the participants experienced low to medium workload on the NASA-TLX [72] (INVRQ: M=18.64, SD=11.38, OUTVRQ: M=14.40, SD=9.42). Paired t-tests revealed significant differences on physical demand (t_{37}=4.14, p<.01) and effort (t_{37}=3.00, p<.01) subscales but not on mental demand (t_{37}=1.42, p>.05), temporal demand (t_{37}=1.15, p>.05), performance (t_{37}=1.48, p>.05) or frustration (t_{37}=1.76, p>.05). The results are depicted in Figure 5b. We contrasted the mean TLX-scores against Schwind et al.'s values (VR: M=33.16, SD=20.96, PC: M=37.77, SD=19.26) [159]. Our data show a significantly lower workload for both corresponding comparisons INVRQ (t_{72}=3.50, p<.01) and OUTVRQ (t_{72}=6.57, p<.01) conditions.

Qualitative Results

We collected relevant statements from the interviews and dynamically generated categories emerging from the material.



(a) Ratings of presence on IPQ [155] subscales general presence for (GP), spacial presence (SP), involvement (INV) and realism (REAL) split by INVRQS and OUTVRQS conditions.

(b) NASA-TLX [72] results on the subscales: mental demand (MD), physical demand (PD), temporal demand (TD), performance (P), effort (E) and frustration (F) split by questionnaire conditions.

Figure 5. Boxplots of presence and workload split by questionnaire type

We indicate how many participants agree with the central statements made by others. Exact quotes were translated by the authors and are labeled with an ID in parentheses.

The majority of the participants (27/38) stated their *VR experience* as "fun" or "enjoyable". 17 found the INVRQS easy to use and the interaction intuitive. 4 participants preferred the directness of input with the VR controllers. However, 16 reported that sometimes the UI did not respond to their input; 4 described this as frustrating. 26 referred to the OUTVRQS as the "common" and "normal" questionnaires. 4 participants reported higher effort using the mouse and 12 participants would have preferred a faster mouse movement in OUTVRQ. Although the participants highlighted the advantages of familiarity using the mouse, we observed that all users instinctively understood the pointing interaction.

17 users criticized changing from VR to physical reality and stated that not having to switch the system feels more fluent and "better integrated" (P12). 4 users addressed the BIP due to the change of medium and how it might affect data quality.

Regarding *presentation*, 31 participants confirmed that readability and font size of the INVRQS were good and the questionnaire canvas was positioned well. P34 disliked not having a full overview of the displayed content in contrast to the notebook screen. 13 users mentioned that blurred edges in the HMD required head movement for reading. 8 users found it strenuous to fill out the questionnaires in VR. 5 of them suggested to provide a chair, P38 sat on the floor. Concerning *data validity*, 4 participants hypothesized that maintaining the immersion is "better for the results" (P11). When answering how they feel, 14 participants found it beneficial to do so in VR, because they still were in the situation: "You don't need to recall how you just felt." (P9). Correspondingly, 8 valued the immediacy of the surveying in VR.

12 stated the INVRQS were entertaining: "I didn't know filling out questionnaires can be fun!" (P33). Accordingly, OUTVRQS were referred to as "dryer" (P4) or boring by 12 users, and P35 stated that a pleasant setting can be motivating to fill out the survey. At the same time, such positive excitement may influence certain measures (e.g. affect).

Finally, we asked the participants how they would like to answer questionnaires if the they would have to repeat the study. A majority of 31/38 would prefer to do it in VR, 5 on a computer and 2 had ambivalent opinions.

Discussion

The high ratings on the IPQ [155] with positive differences against neutral on all subscales but REAL, the ratings of the game and the control over the virtual bow, as well as the qualitative statements indicate that the game provided a high sense of presence. This indicates that our experiment design, which intended to simulate a realistic VR user study scenario, was successful. In alignment with results by Schwind et al. [159], we could not find any differences between the conditions on presence. However, our data show no differences in consistency and we cannot confirm their findings of lower variances of presence when surveying in VR. This supports prior findings that presence questionnaires are inadequate to assess BIPs [171]. Further, literature suggests that presence should be assessed behaviourally [48, 166, 172] or physiologically [24, 39, 116, 170].

With the UMUX, we measured high usability scores for both questionnaire tools (INVRQ: M=77.35, SD=18.35, OUTVRQ: M=86.21, SD=9.4). However, the OUTVRQS were rated higher with a medium effect size (Cohen's d=.60). A possible explanation was given in the interviews: The participants sometimes had to repeatedly click on a UI element of the INVRQ for selection. The UMUX score allows a comparison with the SUS [46]. According to Bangor the scores are in a highly acceptable range [14], discarding the concerns of the experts. Similarly on the TLX, we measured a comparatively low workload [65]. However, the physical demand and effort are significantly higher in the INVRQ condition. As 5 participants stated, this could be attributed to the fact that they were standing and using the VR controller in mid-air rather than sitting on a chair and using the mouse with a resting hand on the desk. The comparable duration is consistent with the TLX ratings for temporal demand and allows to infer that answering questions in VR does not affect the duration of a user study.

Limitations and Future Work

Our research only investigates the design aspects of the anchoring and interaction modality with INVRQS. This first step was essential to establish a convenient INVRQ design that can be applied in VR user research. Due to small sample size and non-diverse participants in the pre-study our design recommendations are not conclusive. However, the high UMUX scores in the user study confirm the prior results. In future work, we aim to systematically examine further design options (e.g. HUDs, freehand interaction) at different degrees of contextualization and embeddedness as previous work showed for games [53, 169] and VR [52, 198].

We only considered a short assessment of an INVRQ (≈6.5 min) at the end of 1 exposure. Future work should examine the effects of long VR exposures and the effect of INVRQS for repeated between-trials measurements. We also did not evaluate open-ended questions which allow the subjects to respond freely, but require sophisticated methods for text-entry in VR. Although such methods exist [92, 129], they are often less accurate and efficient than out-VR settings, especially for untrained users [92]. In future work, we aim to investigate text-entry and oral assessment of open-ended questions in VR as [37], [121] and 4 surveyed experts suggested. The slightly lower UMUX rating suggests room for improvements for the in-VR questionnaire components and the increased TLX physical demand and effort, together with participant comments and behaviour suggest that investigations into "middle-ground" approaches (e.g. remaining in VR but being seated) could lead to more practical solutions.

CONCLUSION

Subjective self-reports are frequently used in VR user studies and administered in the physical domain. This can lead to a break in presence [83], disrupt the immersive experience [91] and bias the responses [159]. Embedded questionnaires in the VE reinforce the association of VR and the subjective responses. Although different presentation methods of the questionnaires may affect the results, contemporary research has no shared agreement or validated assessment methods of self-reports in VR user studies. This work aggregates the contemporary body of research, VR expert perspectives and the user experience of INVRQS.

Our first research question (RQ1) investigates current applications of questionnaires in VR user research. From the literature analysis, we identified 15 instances of INVRQS. These few examples differ substantially in visualization and interaction, emphasizing the lack of validated surveying procedures in VR user research. For a comprehensive understanding of the advantages and challenges of INVRQS (RQ2), we conducted an online survey with 67 VR experts. 43/67 of researchers see the importance of embedding questionnaires directly into the VE. To explore presentation and interaction modalities suitable for INVRQS (RQ3), we conducted 2 user studies, in which we first identified world-anchoring and pointing as most adequate design choices to administer questionnaires in VR and then contrasted an INVRQ against a common screen-based OUTVRQ. Although the results show lower usability and higher physical demand and effort of INVRQS, the ratings are within tolerable range and the majority of participants stated a positive attitude towards INVRQ.

24/67 of the experts rated the usefulness of INVRQS below neutral and raised concerns regarding usability (10) and the required time for answering (5). The high UMUX-scores and comparable completion time between INVRQS and OUTVRQS defy these objections. Moreover, the majority of the participants (31/38) would choose INVRQS over OUTVRQS. This result may be partially attributed to a novelty effect of VR and might weaken with a wider dissemination of VR technology. Similarly, our online survey showed that 29/67 of the experts have already applied INVRQS in user studies and mainly consider them as useful and effective. Based on our findings from the literature review, the survey and the user studies, we advocate that presenting questionnaires in the VE helps the participants to report their experience in a convenient, non-interruptive manner.

Based on the results from the 4 presented studies we conclude: (i) researchers should to apply INVRQS in their user studies, (ii) pointing and world-anchoring are usable ways to realise INVRQS, (iii) participants prefer using INVRQS over OUTVRQS and (iv) researcher should be aware of slightly raised physical and mental demands when using INVRQS.

Similar to the establishment of standardized questionnaires that are empirically validated, we propose moving towards standardizing questionnaire implementation and presentation methods in VR, similar to a quasi-standardization that has already occurred in more traditional screen-based questionnaires due to the prevalence of selected survey tools, such as Google Forms or LimeSurvey [57]. Future work will need to seek for comparative experimental evidence on how the questionnaire modality affects the reliability of the the measurements. As with most design choices, there is no absolute right or wrong. However, researchers should be aware how their measurement methods influence the data. This research lays the groundwork for a design theory of INVRQS to provide validated and standardized methods of question-asking in VR.

ACKNOWLEDGEMENTS

This project was partly funded by the European Unions' Horizon 2020 research and innovation program (No 688244), project first.stage, Klaus Tschira Stiftung gGmbH, and by the German Research Foundation DFG, as part of Collaborative Research Center (Sonderforschungsbereich) 1320 "EASE – Everyday Activity Science and Engineering", University of Bremen (https://ease-crc.org/). The research was conducted in subproject P01 *Embodied semantics for the language of action and change*.

REFERENCES

[1] Ethel Mary Abernethy. 1940. The Effect of Changed Environmental Conditions Upon the Results of College Examinations. *The Journal of Psychology* 10, 2 (Oct.

1940), 293–301. DOI: http://dx.doi.org/10.1080/00223980.1940.9917005

[2] Parastoo Abtahi, Mar Gonzalez-Franco, Eyal Ofek, and Anthony Steed. 2019a. I'm a Giant: Walking in Large Virtual Environments at High Speed Gains. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, Glasgow, Scotland Uk, 522:1–522:13. DOI: http://dx.doi.org/10/gf3kzd

[3] Parastoo Abtahi, Benoit Landry, Jackie (Junrui) Yang, Marco Pavone, Sean Follmer, and James A. Landay. 2019b. Beyond The Force: Using Quadcopters to Appropriate Objects and the Environment for Haptics in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, Glasgow, Scotland Uk, 359:1–359:13. DOI:http://dx.doi.org/10/gf3krq

[4] Lu Ann Aday, Llewellyn J. Cornelius, Foreword Steven, B. Cohen, Lu Ann Aday, Llewellyn J. Cornelius, Foreword Steven, B. Cohen, and A. Wiley Imprint. 2006. *Designing and Conducting Health Surveys: A Comprehensive Guide* (thrid edition ed.). Jossey-Bass Publishers.

[5] G. Ahamba, D. J. Roberts, and P. Eachus. 2017. Triggering and Measuring Social Inhibitory Response in Humans Immersed in Interactions with Virtual Humans. *Annual Review of Cybertherapy and Telemedicine* 15 (April 2017), 185–188.

[6] Majed Al Zayer, Isayas B. Adhanom, Paul MacNeilage, and Eelke Folmer. 2019. The Effect of Field-of-View Restriction on Sex Bias in VR Sickness and Spatial Navigation Performance. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, Glasgow, Scotland Uk, 354:1–354:12. DOI:http://dx.doi.org/10/gf3krr

[7] Judith Amores, Xavier Benavides, and Pattie Maes. 2015. ShowMe: A Remote Collaboration System That Supports Immersive Gestural Communication. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '15)*. ACM, New York, NY, USA, 1343–1348. DOI: http://dx.doi.org/10.1145/2702613.2732927

[8] I. M. Arafat, S. M. Shahnewaz Ferdous, and J. Quarles. 2018. Cybersickness-Provoking Virtual Reality Alters Brain Signals of Persons with Multiple Sclerosis. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 1–120. DOI: http://dx.doi.org/10/gf39qg

[9] Rahul Arora, Rubaiat Habib Kazi, Fraser Anderson, Tovi Grossman, Karan Singh, and George Fitzmaurice. 2017. Experimental Evaluation of Sketching on Surfaces in VR. ACM Press, 5643–5654. DOI: http://dx.doi.org/10.1145/3025453.3025474

[10] Association for Computing Machinery. 2019. ACM Digital Library. https://dl.acm.org/. (2019).

[11] Mahdi Azmandian, Timofey Grechkin, and Evan Suma Rosenberg. 2017. An Evaluation of Strategies for Two-User Redirected Walking in Shared Physical Spaces. IEEE, 91–98. DOI: http://dx.doi.org/10.1109/VR.2017.7892235

[12] Earl Babbie. 1990. *Survey Research Methods* (2nd ed. ed.). Wadsworth, Balmont, Calif.

[13] Chris Baker, Nancy Pistrang, and Robert Elliott. 2002. *Research Methods in Clinical Psychology: An Introduction for Students and Practitioners* (second edition ed.). John Wiley & Sons Ltd.

[14] Aaron Bangor, Kortum, Philip, and Miller, James. 2009. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. *Journal of usability studies* 4, 3 (2009), 114–123.

[15] Soumya C. Barathi, Daniel J. Finnegan, Matthew Farrow, Alexander Whaley, Pippa Heath, Jude Buckley, Peter W. Dowrick, Burkhard C. Wuensche, James L. J. Bilzon, Eamonn O'Neill, and Christof Lutteroth. 2018. Interactive Feedforward for Improving Performance and Maintaining Intrinsic Motivation in VR Exergaming. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 408:1–408:14. DOI:http://dx.doi.org/10.1145/3173574.3173982

[16] Jiwan Bhandari, Sam Tregillus, and Eelke Folmer. 2017. Legomotion: Scalable Walking-Based Virtual Locomotion. ACM Press, 1–8. DOI: http://dx.doi.org/10.1145/3139131.3139133

[17] Anol Bhattacherjee. 2012. *Social Science Research: Principles, Methods, and Practices* (second edition ed.). Number 3 in Textbooks Collection.

[18] Sam B Bhayani and Gerald L Andriole. 2005. Three-Dimensional (3D) Vision: Does It Improve Laparoscopic Skills? An Assessment of a 3D Head-Mounted Visualization System. *Reviews in Urology* 7, 4 (2005), 211–214.

[19] Bobby Bodenheimer, Sarah Creem-Regehr, Jeanine Stefanucci, Elena Shemetova, and William B. Thompson. 2017. Prism Aftereffects for Throwing with a Self-Avatar in an Immersive Virtual Environment. IEEE, 141–147. DOI: http://dx.doi.org/10.1109/VR.2017.7892241

[20] M. Boldt, M. Bonfert, I. Lehne, M. Cahnbley, K. Korschinq, L. Bikas, S. Finke, M. Hanci, V. Kraft, B. Liu, T. Nguyen, A. Panova, R. Singh, A. Steenbergen, R. Malaka, and J. Jan Smeddinck. 2018. You Shall Not Pass: Non-Intrusive Feedback for Virtual Walls in VR Environments with Room-Scale Mapping. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 143–150. DOI: http://dx.doi.org/10.1109/VR.2018.8446177

[21] Stéphane Bouchard, Julie St-Jacques, Geneviève Robillard, and Patrice Renaud. 2008. Anxiety Increases

the Feeling of Presence in Virtual Reality. *Presence: Teleoperators and Virtual Environments* 17, 4 (July 2008), 376–391. DOI:http://dx.doi.org/10/bctvw8

[22] LouAnne E. Boyd, Saumya Gupta, Sagar B. Vikmani, Carlos M. Gutierrez, Junxiang Yang, Erik Linstead, and Gillian R. Hayes. 2018. vrSocial: Toward Immersive Therapeutic VR Systems for Children with Autism. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 204:1–204:12. DOI: http://dx.doi.org/10.1145/3173574.3173778

[23] Norman M. Bradburn and Seymour Sudman. 1979. *Improving Interview Methods and Questionnaire Design - Response Effects to Threatening Qurstions in Survey Research*. Jossey-Bass Publishers.

[24] Andrea Brogni, Vinoba Vinayagamoorthy, Anthony Steed, and Mel Slater. 2006. Variations in Physiological Responses of Participants During Different Stages of an Immersive Virtual Environment Experiment. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST '06)*. ACM, Limassol, Cyprus, 376–382. DOI: http://dx.doi.org/10/b9drht

[25] John Brooke. 1996. SUS - A Quick and Dirty Usability Scale. (1996), 7.

[26] Emily Brown and Paul Cairns. 2004. A Grounded Investigation of Game Immersion. In *Extended Abstracts of the 2004 Conference on Human Factors and Computing Systems - CHI '04*. ACM Press, Vienna, Austria, 1297. DOI: http://dx.doi.org/10.1145/985921.986048

[27] Peter W. S. Butcher, Nigel W. John, and Panagiotis D. Ritsos. 2019. VRIA - A Framework for Immersive Analytics on the Web. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. ACM, Glasgow, Scotland Uk, LBW2615:1–LBW2615:6. DOI: http://dx.doi.org/10/gf3kb8

[28] Zekun Cao, Jason Jerald, and Regis Kopper. 2018. Visually-Induced Motion Sickness Reduction via Static and Dynamic Rest Frames. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 105–112. DOI:http://dx.doi.org/10/gf39qh

[29] Sibo Chen, Junce Wang, Santiago Guerra, Neha Mittal, and Soravis Prakkamakul. 2019. Exploring Word-Gesture Text Entry Techniques in Virtual Reality. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. ACM, Glasgow, Scotland Uk, LBW0233:1–LBW0233:6. DOI: http://dx.doi.org/10/gf3kb4

[30] Alan Cheng, Lei Yang, and Erik Andersen. 2017b. Teaching Language and Culture with a Virtual Reality Game. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 541–549. DOI:http://dx.doi.org/10/gf78wz

[31] Lung-Pan Cheng, Eyal Ofek, Christian Holz, Hrvoje Benko, and Andrew D. Wilson. 2017a. Sparse Haptic Proxy: Touch Feedback in Virtual Environments Using a General Passive Prop. ACM Press, 3718–3728. DOI: http://dx.doi.org/10.1145/3025453.3025753

[32] Bernard C.K. Choi and Anita W.P. Pak. 2004. A Catalog of Biases in Questionnaires. *Preventing Chronic Disease* 2, 1 (Dec. 2004).

[33] Tanvir Irfan Chowdhury, Sharif Mohammad Shahnewaz Ferdous, and John Quarles. 2017. Information Recall in a Virtual Reality Disability Simulation. ACM Press, 1–10. DOI: http://dx.doi.org/10.1145/3139131.3139143

[34] Sebastian Cmentowski, Andrey Krekhov, and Jens Krueger. 2019. Outstanding: A Perspective-Switching Technique for Covering Large Distances in VR Games. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. ACM, Glasgow, Scotland Uk, LBW1612:1–LBW1612:6. DOI: http://dx.doi.org/10/gf3kcd

[35] Mihaly Csikszentmihalyi. 1990. *Flow: The Psychology of Optimal Experience*. Harper & Row, New York.

[36] Lynnda M. Dahlquist, Kristine D. McKenna, Katia K. Jones, Lindsay Dillinger, Karen E. Weiss, and Claire Sonntag Ackerman. 2007. Active and Passive Distraction Using a Head-Mounted Display Helmet: Effects on Cold Pressor Pain in Children. *Health Psychology: Official Journal of the Division of Health Psychology, American Psychological Association* 26, 6 (Nov. 2007), 794–801. DOI: http://dx.doi.org/10/cmp2vx

[37] Donald Degraen, André Zenner, and Antonio Krüger. 2019. Enhancing Texture Perception in Virtual Reality Using 3D-Printed Hair Structures. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, Glasgow, Scotland Uk, 249:1–249:12. DOI: http://dx.doi.org/10/gfz8mf

[38] Arindam Dey, Hao Chen, Mark Billinghurst, and Robert W. Lindeman. 2018. Effects of Manipulating Physiological Feedback in Immersive Virtual Environments. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '18)*. ACM, Melbourne, VIC, Australia, 101–111. DOI:http://dx.doi.org/10/gf3ngn

[39] Cath Dillon, Edmund Keogh, and Jonathan Freeman. 2002. It's Been Emotional: Affect, Physiology, and Presence. In *Proceedings of the Fifth Annual International Workshop on Presence, Porto, Portugal*.

[40] Tilman Dingler, Kai Kunze, and Benjamin Outram. 2018. VR Reading UIs: Assessing Text Parameters for Reading in VR. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (CHI EA '18)*. ACM, New York, NY, USA,

LBW094:1–LBW094:6. DOI: http://dx.doi.org/10.1145/3170427.3188695

[41] Alan Dix. 2009. Human-Computer Interaction. In *Encyclopedia of Database Systems*. Springer, 1327–1331.

[42] E. Ebrahimi, L. S. Hartman, A. Robb, C. C. Pagano, and S. V. Babu. 2018. Investigating the Effects of Anthropomorphic Fidelity of Self-Avatars on Near Field Depth Perception in Immersive Virtual Environments. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 1–8. DOI: http://dx.doi.org/10/gf39qs

[43] Claudia Faita, Federico Vanni, Camilla Tanca, Emanuele Ruffaldi, Marcello Carrozzino, and Massimo Bergamasco. 2016. Investigating the Process of Emotion Recognition in Immersive and Non-Immersive Virtual Technological Setups. ACM Press, 61–64. DOI: http://dx.doi.org/10.1145/2993369.2993395

[44] A. S. Fernandes and S. K. Feiner. 2016. Combating VR Sickness through Subtle Dynamic Field-of-View Modification. In *2016 IEEE Symposium on 3D User Interfaces (3DUI)*. 201–210. DOI: http://dx.doi.org/10/gcmfn6

[45] Andy Field and Graham Hole. 2003. *How to Design and Report Experiments*. Sage Publications Ltd, London ; Thousand Oaks, Calif.

[46] Kraig Finstad. 2010. The Usability Metric for User Experience. *Interacting with Computers* 22, 5 (Sept. 2010), 323–327. DOI:http://dx.doi.org/10/ctr6r4

[47] Daniel Freeman, Polly Haselton, Jason Freeman, Bernhard Spanlang, Sameer Kishore, Emily Albery, Megan Denne, Poppy Brown, Mel Slater, and Alecia Nickless. 2018. Automated Psychological Therapy Using Immersive Virtual Reality for Treatment of Fear of Heights: A Single-Blind, Parallel-Group, Randomised Controlled Trial. *The Lancet Psychiatry* 5, 8 (Aug. 2018), 625–632. DOI: http://dx.doi.org/10/gd2h5n

[48] Jonathan Freeman, S. E. Avons, Ray Meddis, Don E. Pearson, and Wijnand IJsselsteijn. 2000. Using Behavioral Realism to Estimate Presence: A Study of the Utility of Postural Responses to Motion Stimuli. *Presence: Teleoperators and Virtual Environments* 9, 2 (April 2000), 149–164. DOI: http://dx.doi.org/10/drkrr8

[49] Jonathan Freeman, S. E. Avons, Don E. Pearson, and Wijnand A. IJsselsteijn. 1999. Effects of Sensory Information and Prior Experience on Direct Subjective Ratings of Presence. *Presence: Teleoperators and Virtual Environments* 8, 1 (Feb. 1999), 1–13. DOI: http://dx.doi.org/10/dg7fn6

[50] R. Fribourg, F. Argelaguet, L. Hoyet, and A. Lécuyer. 2018. Studying the Sense of Embodiment in VR Shared Experiences. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 273–280. DOI:http://dx.doi.org/10/gf39p6

[51] Thomas Fröhlich, Dmitry Alexandrovsky, Timo Stabbert, Tanja Döring, and Rainer Malaka. 2018. VRBox: A Virtual Reality Augmented Sandbox for Immersive Playfulness, Creativity and Exploration. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '18)*. ACM, Melbourne, VIC, Australia, 153–162. DOI: http://dx.doi.org/10/gfsvjx

[52] Julian Frommel, Kim Fahlbusch, Julia Brich, and Michael Weber. 2017. The Effects of Context-Sensitive Tutorials in Virtual Reality Games. ACM Press, 367–375. DOI: http://dx.doi.org/10.1145/3116595.3116610

[53] Julian Frommel, Michael Weber, Katja Rogers, Julia Brich, Daniel Besserer, Leonard Bradatsch, Isabel Ortinau, Ramona Schabenberger, Valentin Riemer, and Claudia Schrader. 2015. Integrated Questionnaires: Maintaining Presence in Game Environments for Self-Reported Data Acquisition. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15*. ACM Press, London, United Kingdom, 359–368. DOI: http://dx.doi.org/10/gfvjmx

[54] Wei Gai, Cheng Lin, Chenglei Yang, Yulong Bian, Chia Shen, Xiangxu Meng, Lu Wang, Juan Liu, Mingda Dong, and Chengjie Niu. 2017. Supporting Easy Physical-to-Virtual Creation of Mobile VR Maze Games: A New Genre. ACM Press, 5016–5028. DOI: http://dx.doi.org/10.1145/3025453.3025494

[55] D. Gall and M. E. Latoschik. 2018. The Effect of Haptic Prediction Accuracy on Presence. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 73–80. DOI:http://dx.doi.org/10/gf39qm

[56] Maryrose Gerardi, Barbara Olasov Rothbaum, Kerry Ressler, Mary Heekin, and Albert Rizzo. 2008. Virtual reality exposure therapy using a virtual Iraq: Case report. *Journal of Traumatic Stress* 21, 2 (April 2008), 209–213. DOI:http://dx.doi.org/10/djjmn5

[57] Limesurvey GmbH. 2003. LimeSurvey: An Open Source Survey Tool. (2003).

[58] D. R. Godden and A. D. Baddeley. 1975. Context-Dependent MemoryIn Two Natural Environments: On Land and Underwater. *British Journal of Psychology* 66, 3 (Aug. 1975), 325–331. DOI:http://dx.doi.org/10/cjgxhp

[59] David Goedicke, Jamy Li, Vanessa Evers, and Wendy Ju. 2018. VR-OOM: Virtual Reality On-rOad Driving siMulation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 165:1–165:11. DOI:http://dx.doi.org/10.1145/3173574.3173739

[60] Jeffrey I. Gold, Seok Hyeon Kim, Alexis J. Kant, Michael H. Joseph, and Albert "Skip" Rizzo. 2006. Effectiveness of Virtual Reality for Pediatric Pain Distraction during IV Placement. *CyberPsychology & Behavior* 9, 2 (April 2006), 207–212. DOI: http://dx.doi.org/10/fwsrpx

[61] Google. 2018. Rendering Text in WebVR | Web. https://developers.google.com/web/showcase/2017/within. (Sept. 2018).

[62] Google Developers. 2017. Designing Screen Interfaces for VR (Google I/O '17). (2017).

[63] Antti Granqvist, Tapio Takala, Jari Takatalo, and Perttu Hämäläinen. 2018. Exaggeration of Avatar Flexibility in Virtual Reality. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '18)*. ACM, Melbourne, VIC, Australia, 201–209. DOI:http://dx.doi.org/10/gfvpff

[64] Anthony G. Greenwald and Brian A. Nosek. 2001. Health of the Implicit Association Test at Age 3. *Experimental Psychology* 48, 2 (April 2001), 85–93. DOI:http://dx.doi.org/10/c4k2qx

[65] Rebecca A. Grier. 2015. How High Is High? A Meta-Analysis of NASA-TLX Global Workload Scores. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 59, 1 (Sept. 2015), 1727–1731. DOI:http://dx.doi.org/10/gf447w

[66] Nathan Navarro Griffin, James Liu, and Eelke Folmer. 2018. Evaluation of Handsbusy vs Handsfree Virtual Locomotion. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '18)*. ACM, Melbourne, VIC, Australia, 211–219. DOI:http://dx.doi.org/10/gf3ngm

[67] J. Grubert, L. Witzani, E. Ofek, M. Pahud, M. Kranz, and P. O. Kristensson. 2018a. Effects of Hand Representations for Typing in Virtual Reality. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 151–158. DOI: http://dx.doi.org/10/gf39qd

[68] J. Grubert, L. Witzani, E. Ofek, M. Pahud, M. Kranz, and P. O. Kristensson. 2018b. Text Entry in Immersive Head-Mounted Display-Based Virtual Reality Using Standard Keyboards. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 159–166. DOI:http://dx.doi.org/10/gf39qc

[69] Jan Gugenheimer, Evgeny Stemasov, Julian Frommel, and Enrico Rukzio. 2017. ShareVR: Enabling Co-Located Experiences for Virtual Reality between HMD and Non-HMD Users. ACM Press, 4021–4033. DOI:http://dx.doi.org/10.1145/3025453.3025683

[70] Jan Gugenheimer, Evgeny Stemasov, Harpreet Sareen, and Enrico Rukzio. 2018. FaceDisplay: Towards Asymmetric Multi-User Interaction for Nomadic Virtual Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 54:1–54:13. DOI:http://dx.doi.org/10.1145/3173574.3173628

[71] John Paulin Hansen, Astrid Kofod Trudslev, Sara Amdi Harild, Alexandre Alapetite, and Katsumi Minakata. 2019. Providing Access to VR Through a Wheelchair. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. ACM, Glasgow, Scotland Uk, CS07:1–CS07:8. DOI: http://dx.doi.org/10/gf3kb2

[72] Sandra G. Hart. 1986. *NASA Task Load Index (TLX). Volume 1.0; Paper and Pencil Package*. Technical Report.

[73] Jeremy Hartmann, Christian Holz, Eyal Ofek, and Andrew D. Wilson. 2019. RealityCheck: Blending Virtual Environments with Situated Physical Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, Glasgow, Scotland Uk, 347:1–347:12. DOI: http://dx.doi.org/10/gf2hj4

[74] Tianyu He, Xiaoming Chen, Zhibo Chen, Ye Li, Sen Liu, Junhui Hou, and Ying He. 2017. Immersive and Collaborative Taichi Motion Learning in Various VR Environments. IEEE, 307–308. DOI: http://dx.doi.org/10.1109/VR.2017.7892299

[75] Teresa Hirzle, Jan Rixen, Jan Gugenheimer, and Enrico Rukzio. 2018. WatchVR: Exploring the Usage of a Smartwatch for Interaction in Mobile Virtual Reality. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (CHI EA '18)*. ACM, New York, NY, USA, LBW634:1–LBW634:6. DOI:http://dx.doi.org/10.1145/3170427.3188629

[76] Philipp Hock, Sebastian Benedikter, Jan Gugenheimer, and Enrico Rukzio. 2017. CarVR: Enabling In-Car Virtual Reality Entertainment. ACM Press, 4034–4044. DOI:http://dx.doi.org/10.1145/3025453.3025665

[77] Matthias Hoppe, Jakob Karolus, Felix Dietz, PawełW. Woźniak, Albrecht Schmidt, and Tonja-Katrin Machulla. 2019. VRsneaky: Increasing Presence in VR Through Gait-Aware Auditory Feedback. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, Glasgow, Scotland Uk, 546:1–546:9. DOI: http://dx.doi.org/10/gf3kzc

[78] HTC. 2016. VIVE Now Shipping Immediately From HTC, Retail Partners Expand Demo Locations. https://www.htc.com/us/newsroom/2016-06-07/. (2016).

[79] Brent E. Insko. 2003. Measuring Presence: Subjective, Behavioral and Physiological Methods. In *Being There: Concepts, Effects and Measurements of User Presence in Synthetic Environments*. IOS Press, Amsterdam, Netherlands, 109–119.

[80] Institute of Electrical and Electronics Engineers. 2019. IEEE Xplore Digital Library. https://ieeexplore.ieee.org/. (2019).

[81] International Organization for Standardization (ISO). 2006. Ergonomics of Human-System Interaction — Part 110: Dialogue Principles (ISO 9241-110). (2006).

[82] Charlene Jennett, Anna L. Cox, Paul Cairns, Samira Dhoparee, Andrew Epps, Tim Tijs, and Alison Walton. 2008. Measuring and Defining the Experience of Immersion in Games. *International Journal of Human-Computer Studies* 66, 9 (Sept. 2008), 641–661. DOI:http://dx.doi.org/10/ct9vk5

[83] Jason Jerald. 2016. *The VR Book: Human-Centered Design for Virtual Reality* (first edition ed.). Number 8 in ACM Books. acm, Association for Computing Machinery, New York. OCLC: 945367096.

[84] S. Jung, P. J. Wisniewski, and C. E. Hughes. 2018. In Limbo: The Effect of Gradual Visual Transition Between Real and Virtual on Virtual Body Ownership Illusion and Presence. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 267–272. DOI:http://dx.doi.org/10/gf39p7

[85] HyeongYeop Kang, Geonsun Lee, Seongsu Kwon, Ohung Kwon, Seongpil Kim, and JungHyun Han. 2018. Flotation Simulation in a Cable-Driven Virtual Environment – A Study with Parasailing. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 632:1–632:11. DOI: http://dx.doi.org/10.1145/3173574.3174206

[86] Oliver Beren Kaul and Michael Rohs. 2017. HapticHead: A Spherical Vibrotactile Grid around the Head for 3D Guidance in Virtual and Augmented Reality. ACM Press, 3729–3740. DOI: http://dx.doi.org/10.1145/3025453.3025684

[87] Robert S. Kennedy, Norman E. Lane, Kevin S. Berbaum, and Michael G. Lilienthal. 1993. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology* 3, 3 (July 1993), 203–220. DOI:http://dx.doi.org/10/bbh54v

[88] Behrang Keshavarz and Heiko Hecht. 2011. Validating an Efficient Method to Quantify Motion Sickness. *Human Factors* 53, 4 (Aug. 2011), 415–426. DOI: http://dx.doi.org/10/dz78cg

[89] Youngwon Ryan Kim and Gerard J. Kim. 2017. Presence and Immersion of "Easy" Mobile VR with Open Flip-on Lenses. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology (VRST '17)*. ACM, New York, NY, USA, 38:1–38:7. DOI:http://dx.doi.org/10.1145/3139131.3139147

[90] Tina Kjær, Christoffer B. Lillelund, Mie Moth-Poulsen, Niels C. Nilsson, Rolf Nordahl, and Stefania Serafin. 2017. Can You Cut It?: An Exploration of the Effects of Editing in Cinematic Virtual Reality. ACM Press, 1–4. DOI:http://dx.doi.org/10.1145/3139131.3139166

[91] Jarrod Knibbe, Jonas Schjerlund, Mathias Petraeus, and Kasper Hornb\{ae}k. 2018. The Dream Is Collapsing: The Experience of Exiting VR. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 483:1–483:13. DOI: http://dx.doi.org/10.1145/3173574.3174057

[92] Pascal Knierim, Valentin Schwind, Anna Maria Feit, Florian Nieuwenhuizen, and Niels Henze. 2018. Physical Keyboards in Virtual Reality: Analysis of Typing Performance and Effects of Avatar Hands. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 345:1–345:9. DOI: http://dx.doi.org/10.1145/3173574.3173919

[93] Julian Kreimeier and Timo Götzelmann. 2019. First Steps Towards Walk-In-Place Locomotion and Haptic Feedback in Virtual Reality for Visually Impaired. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. ACM, Glasgow, Scotland Uk, LBW2214:1–LBW2214:6. DOI: http://dx.doi.org/10/gf3kcb

[94] Andrey Krekhov, Sebastian Cmentowski, Katharina Emmerich, Maic Masuch, and Jens Krüger. 2018. GulliVR: A Walking-Oriented Technique for Navigation in Virtual Reality Games Based on Virtual Body Resizing. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '18)*. ACM, Melbourne, VIC, Australia, 243–256. DOI:http://dx.doi.org/10/gf3ngk

[95] Andrey Krekhov, Katharina Emmerich, Philipp Bergmann, Sebastian Cmentowski, and Jens Krüger. 2017. Self-Transforming Controllers for Virtual Reality First Person Shooters. ACM Press, 517–529. DOI: http://dx.doi.org/10.1145/3116595.3116615

[96] Heinz W. Krohne, Boris Egloff, Carl-Walter Kohlmann, and Anja Tausch. 1996. Untersuchungen Mit Einer Deutschen Version Der "Positive and Negative Affect Schedule"(PANAS). *Diagnostica-Gottingen-* 42 (1996), 139–156.

[97] W. Kruger, C. Bohn, B. Frohlich, H. Schuth, W. Strauss, and G. Wesche. 1995. The Responsive Workbench: A Virtual Work Environment. *Computer* 28, 7 (July 1995), 42–48. DOI: http://dx.doi.org/10/cvbntz

[98] D. M. Krum, S. Kang, and T. Phan. 2018. Influences on the Elicitation of Interpersonal Space with Virtual Humans. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 223–9. DOI: http://dx.doi.org/10/gf39p8

[99] David M. Krum, Sin-Hwa Kang, Thai Phan, Lauren Cairco Dukes, and Mark Bolas. 2017. Social Impact of Enhanced Gaze Presentation Using Head Mounted Projection. In *Proceedings of the Human-Computer Interaction International Conference*. Springer International Publishing.

[100] L. Kruse, E. Langbehn, and F. Stelnlcke. 2018. I Can See on My Feet While Walking: Sensitivity to Translation Gains with Visible Feet. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 305–312. DOI:http://dx.doi.org/10/gf39p4

[101] Jérémy Lacoche, Nico Pallamin, Thomas Boggini, and Jérôme Royan. 2017. Collaborators Awareness for User Cohabitation in Co-Located Collaborative Virtual Environments. ACM Press, 1–9. DOI: http://dx.doi.org/10.1145/3139131.3139142

[102] Joseph J. LaViola, Ernst Kruijff, Doug A. Bowman, Ryan P. McMahan, and Ivan Poupyrev. 2017. *3D User Interfaces: Theory and Practice*. Addison-Wesley.

[103] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. 2017. *Research Methods in Human-Computer Interaction* (2 ed.). Morgan Kaufmann, Cambridge, MA.

[104] Kangdon Lee. 2012. Augmented Reality in Education and Training. *TechTrends* 56, 2 (March 2012), 13–21. DOI:http://dx.doi.org/10/gdz3kh

[105] Myungho Lee, Gerd Bruder, and Gregory F. Welch. 2017. Exploring the Effect of Vibrotactile Feedback through the Floor on Social Presence in an Immersive Virtual Environment. IEEE, 105–111. DOI: http://dx.doi.org/10.1109/VR.2017.7892237

[106] Meng-Yun Liao, Ching-Ying Sung, Hao-Chuan Wang, Wen-Chieh Lin, and Fu-Yin Cherng. 2019. Embodying Historical Learners' Messages As Learning Companions in a VR Classroom. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. ACM, Glasgow, Scotland Uk, LBW1115:1–LBW1115:6. DOI: http://dx.doi.org/10/gf3kcg

[107] Pedro Lopes, Sijing You, Lung-Pan Cheng, Sebastian Marwecki, and Patrick Baudisch. 2017. Providing Haptics to Walls & Heavy Objects in Virtual Reality by Means of Electrical Muscle Stimulation. ACM Press, 1471–1482. DOI: http://dx.doi.org/10.1145/3025453.3025600

[108] Sarah Lopez, Yi Yang, Kevin Beltran, Soo Jung Kim, Jennifer Cruz Hernandez, Chelsy Simran, Bingkun Yang, and Beste F. Yuksel. 2019. Investigating Implicit Gender Bias and Embodiment of White Males in Virtual Reality with Full Body Visuomotor Synchrony. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 557. DOI: http://dx.doi.org/10/gf3kzb

[109] C. Di Loreto, J. Chardonnet, J. Ryard, and A. Rousseau. 2018. WoaH: A Virtual Reality Work-at-Height Simulator. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 281–288. DOI:http://dx.doi.org/10/gf39p5

[110] J. Lugrin, M. Ertl, P. Krop, R. Klüpfel, S. Stierstorfer, B. Weisz, M. Rück, J. Schmitt, N. Schmidt, and M. E. Latoschik. 2018. Any "Body" There? Avatar Visibility Effects in a Virtual Reality Game. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 17–24. DOI:http://dx.doi.org/10/gf39qq

[111] Raquel Macedo, Nuno Correia, and Teresa Romão. 2019. Paralympic VR Game: Immersive Game Using Virtual Reality and Video. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. ACM, Glasgow, Scotland Uk, LBW0235:1–LBW0235:6. DOI: http://dx.doi.org/10/gf3kb3

[112] I. Scott MacKenzie, Tatu Kauppinen, and Miika Silfverberg. 2001. Accuracy Measures for Evaluating Computer Pointing Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '01)*. ACM, Seattle, Washington, USA, 9–16. DOI:http://dx.doi.org/10.1145/365024.365028 MacKenzie2001.

[113] I. Scott MacKenzie, Abigail Sellen, and William A. S. Buxton. 1991. A Comparison of Input Devices in Element Pointing and Dragging Tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '91)*. ACM, New York, NY, USA, 161–166. DOI: http://dx.doi.org/10.1145/108844.108868

[114] Elhassan Makled, Yomna Abdelrahman, Noha Mokhtar, Valentin Schwind, Slim Abdennadher, and Albrecht Schmidt. 2018. I Like to Move It: Investigating the Effect of Head and Body Movement of Avatars in VR on User's Perception. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (CHI EA '18)*. ACM, New York, NY, USA, LBW099:1–LBW099:6. DOI: http://dx.doi.org/10.1145/3170427.3188573

[115] Willie McClinton, Derek Caprio, Denis Laesker, Blanche Pinto, Sarah Garcia, and Marvin Andujar. 2019. P300-Based 3D Brain Painting in Virtual Reality. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. ACM, Glasgow, Scotland Uk, LBW1119:1–LBW1119:6. DOI: http://dx.doi.org/10/gf3kcf

[116] Michael Meehan. 2001. *Physiological Reaction as an Objective Measure of Presence in Virtual Environments*. Ph.D. Dissertation.

[117] Daniel Mendes, Maurício Sousa, Rodrigo Lorena, Alfredo Ferreira, and Joaquim Jorge. 2017. Using Custom Transformation Axes for Mid-Air Manipulation of 3D Virtual Objects. ACM Press, 1–8. DOI:http://dx.doi.org/10.1145/3139131.3139157

[118] Mark R Mine. 1997. ISAAC: A Meta-Cad System for Virtual Environments. *Computer-Aided Design* 29, 8 (Aug. 1997), 547–553. DOI: http://dx.doi.org/10/b6gzzj

[119] Thomas B. Moeslund, Moritz Störring, and Erik Granum. 2002. A Natural Interface to a Virtual Environment through Computer Vision-Estimated Pointing Gestures. In *Gesture and Sign Language in Human-Computer Interaction (Lecture Notes in Computer Science)*, Ipke Wachsmuth and Timo Sowa (Eds.). Springer Berlin Heidelberg, 59–63.

[120] Aske Mottelson and Kasper Hornbæk. 2017. Virtual Reality Studies Outside the Laboratory. ACM Press, 1–10. DOI:http://dx.doi.org/10.1145/3139131.3139141

[121] C. Mousas. 2018. Performance-Driven Dance Motion Control of a Virtual Partner Character. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 57–64. DOI:http://dx.doi.org/10/gf39qp

[122] Thomas Muender, Anke V. Reinschluessel, Sean Drewes, Dirk Wenig, Tanja Döring, and Rainer Malaka. 2019. Does It Feel Real?: Using Tangibles with Different Fidelities to Build and Explore Scenes in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, Glasgow, Scotland Uk, 673:1–673:12. DOI:http://dx.doi.org/10/gf2d3x

[123] Marcus R. Munafò, Brian A. Nosek, Dorothy V. M. Bishop, Katherine S. Button, Christopher D. Chambers, Nathalie Percie du Sert, Uri Simonsohn, Eric-Jan Wagenmakers, Jennifer J. Ware, and John P. A. Ioannidis. 2017. A Manifesto for Reproducible Science. *Nature Human Behaviour* 1, 1 (Jan. 2017), 0021. DOI:http://dx.doi.org/10/bw28

[124] S. Narang, A. Best, and D. Manocha. 2018. Simulating Movement Interactions Between Avatars Agents in Virtual Worlds Using Human Motion Constraints. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 9–16. DOI:http://dx.doi.org/10/gf39qr

[125] Cuong Nguyen, Stephen DiVerdi, Aaron Hertzmann, and Feng Liu. 2017. Vremiere: In-Headset Virtual Reality Video Editing. ACM Press, 5428–5438. DOI:http://dx.doi.org/10.1145/3025453.3025675

[126] Cuong Nguyen, Stephen DiVerdi, Aaron Hertzmann, and Feng Liu. 2018. Depth Conflict Reduction for Stereo VR Video Interfaces. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 64:1–64:9. DOI:http://dx.doi.org/10.1145/3173574.3173638

[127] Sebastian Oberdörfer, David Heidrich, and Marc Erich Latoschik. 2019. Usability of Gamified Knowledge Learning in VR and Desktop-3D. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, Glasgow, Scotland Uk, 175:1–175:13. DOI:http://dx.doi.org/10/gf3k86

[128] Oculus. 2019. Introduction to Best Practices. https://developer.oculus.com/design/latest/concepts/book-bp/. (Sept. 2019).

[129] Eyal Ofek and Michel Pahud. 2018. Typing in the Virtual Office Environment: Examining the Effects of Keyboard and Hand Representation in VR. (Feb. 2018).

[130] Roshan Lalintha Peiris, Wei Peng, Zikun Chen, Liwei Chan, and Kouta Minamizawa. 2017. ThermoVR: Exploring Integrated Thermal Haptic Feedback with Head Mounted Displays. ACM Press, 5452–5456. DOI:http://dx.doi.org/10.1145/3025453.3025824

[131] Xiaolan Peng, Jin Huang, Linghan Li, Chen Gao, Hui Chen, Feng Tian, and Hongan Wang. 2019. Beyond Horror and Fear: Exploring Player Experience Invoked by Emotional Challenge in VR Games. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. ACM, Glasgow, Scotland Uk, LBW1616:1–LBW1616:6. DOI:http://dx.doi.org/10/gf3kcc

[132] Max Pfeiffer, Tobias Kröger, Jens Seifert, Sulaxan Somaskantharajan, Lukas Jahnich, Tobias Steinblum, Jan Speckamp, and Samuel Navas Medrano. 2019. WONDER – Enhancing VR Training with Electrical Muscle Stimulation. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. ACM, Glasgow, Scotland Uk, LBW0232:1–LBW0232:6. DOI:http://dx.doi.org/10/gf3kb5

[133] Ken Pfeuffer, Matthias J. Geiger, Sarah Prange, Lukas Mecke, Daniel Buschek, and Florian Alt. 2019. Behavioural Biometrics in VR: Identifying People from Body Motion and Relations in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, Glasgow, Scotland Uk, 110:1–110:12. DOI:http://dx.doi.org/10/gf3kx6

[134] Dario Pittera, Elia Gatti, and Marianna Obrist. 2019. I'm Sensing in the Rain: Spatial Incongruity in Visual-Tactile Mid-Air Stimulation Can Elicit Ownership in VR Users. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, Glasgow, Scotland Uk, 132:1–132:15. DOI:http://dx.doi.org/10/gf3kx5

[135] Thammathip Piumsomboon, Gun A. Lee, Jonathon D. Hart, Barrett Ens, Robert W. Lindeman, Bruce H. Thomas, and Mark Billinghurst. 2018. Mini-Me: An Adaptive Avatar for Mixed Reality Remote Collaboration. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 46:1–46:13. DOI:http://dx.doi.org/10.1145/3173574.3173620

[136] Rüdiger Pohl and Rüdiger F. Pohl. 2004. *Cognitive Illusions: A Handbook on Fallacies and Biases in Thinking, Judgement and Memory*. Psychology Press.

[137] Susanne Putze, Dmitry Alexandrovsky, Felix Putze, Sebastian Höffner, Jan D. Smeddinck, and Rainer Malaka. 2020. Breaking The Experience: Effects of Questionnaires in VR User Studies. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Honolulu, HI, USA. DOI:http://dx.doi.org/10.1145/3313831.3376144

[138] J. T. Quaglia and A. Holecek. 2018. Lucid Virtual Dreaming: Antecedents and Consequents of Virtual Lucidity During Virtual Threat. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 1–72. DOI:http://dx.doi.org/10/gf39qn

[139] Nimesha Ranasinghe, Pravar Jain, Shienny Karwita, David Tolley, and Ellen Yi-Luen Do. 2017. Ambiotherm: Enhancing Sense of Presence in Virtual Reality by Simulating Real-World Environmental Conditions. ACM Press, 1731–1742. DOI: http://dx.doi.org/10.1145/3025453.3025723

[140] Michael Rietzler, Florian Geiselhart, Julian Frommel, and Enrico Rukzio. 2018. Conveying the Perception of Kinesthetic Feedback in Virtual Reality Using State-of-the-Art Hardware. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 460:1–460:13. DOI: http://dx.doi.org/10.1145/3173574.3174034

[141] Michael Rietzler, Florian Geiselhart, and Enrico Rukzio. 2017. The Matrix Has You: Realizing Slow Motion in Full-Body Virtual Reality. ACM Press, 1–10. DOI:http://dx.doi.org/10.1145/3139131.3139145

[142] Katja Rogers, Jana Funke, Julian Frommel, Sven Stamm, and Michael Weber. 2019. Exploring Interaction Fidelity in Virtual Reality: Object Manipulation and Whole-Body Movements. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, Glasgow, Scotland Uk, 414:1–414:14. DOI: http://dx.doi.org/10/gf2hj3

[143] Katja Rogers, Giovanni Ribeiro, Rina R. Wehbe, Michael Weber, and Lennart E. Nacke. 2018. Vanishing Importance: Studying Immersive Effects of Game Audio Perception on Player Experiences in Virtual Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 328:1–328:13. DOI:http://dx.doi.org/10.1145/3173574.3173902

[144] Robert Rosenthal and Ralph L. Rosnow. 1984. *Essentials of Behavioral Research: Methods and Data Analysis*. McGraw-Hill, Inc.

[145] D. Roth, C. Klelnbeck, T. Feigl, C. Mutschler, and M. E. Latoschik. 2018. Beyond Replication: Augmenting Social Behaviors in Multi-User Virtual Realities. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 215–222. DOI: http://dx.doi.org/10/gf39p9

[146] Barbara Olasov Rothbaum, Larry Hodges, Samantha Smith, Jeong Hwan Lee, and Larry Price. 2000. A Controlled Study of Virtual Reality Exposure Therapy for the Fear of Flying. *Journal of Consulting and Clinical Psychology* 68, 6 (2000), 1020–1026. DOI: http://dx.doi.org/10/dshf4v

[147] Roy A. Ruddle, Stephen J. Payne, and Dylan M. Jones. 1999. Navigating Large-Scale Virtual Environments: What Differences Occur Between Helmet-Mounted and Desk-Top Displays? *Presence: Teleoperators and Virtual Environments* 8, 2 (April 1999), 157–168. DOI: http://dx.doi.org/10/dg4s2q

[148] Majed Samad, Elia Gatti, Anne Hermes, Hrvoje Benko, and Cesare Parise. 2019. Pseudo-Haptic Weight: Changing the Perceived Weight of Virtual Objects By Manipulating Control-Display Ratio. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, Glasgow, Scotland Uk, 320:1–320:13. DOI: http://dx.doi.org/10/gf2hjv

[149] James B. Sampson. 1993. Cognitive Performance of Individuals Using a Head—Mounted Display While Walking. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 37, 4 (Oct. 1993), 338–342. DOI:http://dx.doi.org/10/gd8cg6

[150] Maria V. Sanchez-Vives and Mel Slater. 2005. From Presence to Consciousness through Virtual Reality. *Nature Reviews Neuroscience* 6, 4 (April 2005), 332. DOI:http://dx.doi.org/10/bh8pn6

[151] Shyam Prathish Sargunam, Kasra Rahimi Moghadam, Mohamed Suhail, and Eric D. Ragan. 2017. Guided Head Rotation and Amplified Head Rotation: Evaluating Semi-Natural Travel and Viewing Techniques in Virtual Reality. IEEE, 19–28. DOI: http://dx.doi.org/10.1109/VR.2017.7892227

[152] Willem E. Saris and Irmtraud N. Gallhofer. 2014. *Design, Evaluation, and Analysis of Questionnaires for Survey Research*. John Wiley & Sons.

[153] Samuel B. Schorr and Allison M. Okamura. 2017. Fingertip Tactile Devices for Virtual Object Manipulation and Exploration. ACM Press, 3115–3119. DOI:http://dx.doi.org/10.1145/3025453.3025744

[154] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. 1999. Embodied Presence in Virtual Environments. In *Visual Representations and Interpretations*, Ray Paton and Irene Neilson (Eds.). Springer London, 269–278.

[155] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. 2001. The Experience of Presence: Factor Analytic Insights. *Presence: Teleoper. Virtual Environ.* 10, 3 (June 2001), 266–281. DOI: http://dx.doi.org/10/cvt2nz

[156] Peter Schulz, Dmitry Alexandrovsky, Felix Putze, Rainer Malaka, and Johannes Schöning. 2019. The

Role of Physical Props in VR Climbing Environments. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, Glasgow, Scotland Uk, 183:1–183:13. DOI: http://dx.doi.org/10/gf3kx3

[157] Valentin Schwind. 2018. Implications of the Uncanny Valley of Avatars and Virtual Characters for Human-Computer Interaction. (2018). DOI: http://dx.doi.org/10/gd6b7h

[158] Valentin Schwind, Pascal Knierim, Lewis Chuang, and Niels Henze. 2017. "Where's Pinky?": The Effects of a Reduced Number of Fingers in Virtual Reality. ACM Press, 507–515. DOI: http://dx.doi.org/10.1145/3116595.3116596

[159] Valentin Schwind, Pascal Knierim, Nico Haas, and Niels Henze. 2019. Using Presence Questionnaires in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, Glasgow, Scotland Uk, 360:1–360:12. DOI:http://dx.doi.org/10/gfz8sr

[160] Valentin Schwind, Pascal Knierim, Cagri Tasci, Patrick Franczak, Nico Haas, and Niels Henze. 2017. "These Are Not My Hands!": Effect of Gender on the Perception of Avatar Hands in Virtual Reality. ACM Press, 1577–1582. DOI: http://dx.doi.org/10.1145/3025453.3025602

[161] Valentin Schwind, Sven Mayer, Alexandre Comeau-Vermeersch, Robin Schweigert, and Niels Henze. 2018. Up to the Finger Tip: The Effect of Avatars on Mid-Air Pointing Accuracy in Virtual Reality. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '18)*. ACM, Melbourne, VIC, Australia, 477–488. DOI: http://dx.doi.org/10/gf3ngj

[162] Sven Seele, Sebastian Misztal, Helmut Buhler, Rainer Herpers, and Jonas Schild. 2017. Here's Looking At You Anyway!: How Important Is Realistic Gaze Behavior in Co-Located Social Virtual Reality Games? ACM Press, 531–540. DOI: http://dx.doi.org/10.1145/3116595.3116619

[163] John J. Shaughnessy and Eugene B. Zechmeister. 1985. *Research Methods in Psychology*. Alfred A. Knopf, New York, NY, US.

[164] Emily Shaw, Tessa Roper, Tommy Nilsson, Glyn Lawson, Sue V.G. Cobb, and Daniel Miller. 2019. The Heat Is On: Exploring User Behaviour in a Multisensory Virtual Environment for Fire Evacuation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, Glasgow, Scotland Uk, 626:1–626:13. DOI: http://dx.doi.org/10/gf3kx9

[165] Lindsay Alexander Shaw, Burkhard Claus WÃŒnsche, Christof Lutteroth, S. Marks, and Rodolphe Callies. 2015. Challenges in Virtual Reality Exergame Design. 162 (Jan. 2015).

[166] Thomas B. Sheridan. 1992. Musings on Telepresence and Virtual Presence. *Presence: Teleoperators and Virtual Environments* 1, 1 (Jan. 1992), 120–126. DOI: http://dx.doi.org/10/gdcftg

[167] Jotaro Shigeyama, Takeru Hashimoto, Shigeo Yoshida, Takuji Narumi, Tomohiro Tanikawa, and Michitaka Hirose. 2019. Transcalibur: A Weight Shifting Virtual Reality Controller for 2D Shape Rendering Based on Computational Perception Model. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, Glasgow, Scotland Uk, 11:1–11:11. DOI: http://dx.doi.org/10/gfz8qr

[168] Ben Shneiderman, Catherine Plaisant, Maxine Cohen, Steven Jacobs, Niklas Elmqvist, and Nicholas Diakopoulos. 2016. *Designing the User Interface: Strategies for Effective Human-Computer Interaction* (6 edition ed.). Pearson, Boston.

[169] Valerie J. Shute. 2011. Stealth Assessment in Computer-Based Games to Support Learning. *Computer games and instruction* 55, 2 (2011), 503–524.

[170] Mel Slater. 2009. Place Illusion and Plausibility Can Lead to Realistic Behaviour in Immersive Virtual Environments. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364, 1535 (Dec. 2009), 3549–3557. DOI:http://dx.doi.org/10/df44xc

[171] Mel Slater and Anthony Steed. 2000. A Virtual Presence Counter. *Presence: Teleoperators and Virtual Environments* 9, 5 (2000), 413–434. DOI: http://dx.doi.org/10/cd5kxg

[172] Mel Slater, Martin Usoh, and Yiorgos Chrysanthou. 1995. The Influence of Dynamic Shadows on Presence in Immersive Virtual Environments. In *Virtual Environments '95 (Eurographics)*, Martin Göbel (Ed.). Springer Vienna, 8–21.

[173] Jeongmin Son, Sunggeun Ahn, Sunbum Kim, and Geehyuk Lee. 2019. Improving Two-Thumb Touchpad Typing in Virtual Reality. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. ACM, Glasgow, Scotland Uk, LBW2217:1–LBW2217:6. DOI: http://dx.doi.org/10/gf3kb9

[174] Maurício Sousa, Daniel Mendes, Soraia Paulo, Nuno Matela, Joaquim Jorge, and Daniel Simões Lopes. 2017. VRRRRoom: Virtual Reality for Radiologists in the Reading Room. ACM Press, 4057–4062. DOI: http://dx.doi.org/10.1145/3025453.3025566

[175] Beatriz Sousa Santos, Paulo Dias, Angela Pimentel, Jan-Willem Baggerman, Carlos Ferreira, Samuel Silva, and Joaquim Madeira. 2008. Head-Mounted Display versus Desktop for 3D Navigation in Virtual Reality: A User Study. *Multimedia Tools and Applications* 41, 1 (Aug. 2008), 161. DOI:http://dx.doi.org/10/ccq4dq

[176] Marco Speicher, Jan Ehrlich, Vito Gentile, Donald Degraen, Salvatore Sorce, and Antonio Krüger. 2019. Pseudo-Haptic Controls for Mid-Air Finger-Based Menu Interaction. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. ACM, Glasgow, Scotland Uk, LBW0285:1–LBW0285:6. DOI: http://dx.doi.org/10/gf3kcj

[177] Marco Speicher, Anna Maria Feit, Pascal Ziegler, and Antonio Krüger. 2018. Selection-Based Text Entry in Virtual Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 647:1–647:13. DOI:http://dx.doi.org/10.1145/3173574.3174221

[178] Misha Sra, Xuhai Xu, and Pattie Maes. 2018. BreathVR: Leveraging Breathing As a Directly Controlled Interface for Virtual Reality Games. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 340:1–340:12. DOI: http://dx.doi.org/10.1145/3173574.3173914

[179] J. Stauffert, F. Niebling, and M. E. Latoschik. 2018. Effects of Latency Jitter on Simulator Sickness in a Search Task. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 121–127. DOI: http://dx.doi.org/10/gf39qf

[180] Carolin Straßmann, Sabrina C. Eimler, Alexander Arntz, Dustin Keßler, Sarah Zielinski, Gabriel Brandenberg, Vanessa Dümpel, and Uwe Handmann. 2019. Relax Yourself - Using Virtual Reality to Enhance Employees' Mental Health and Work Performance. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. ACM, Glasgow, Scotland Uk, LBW0286:1–LBW0286:6. DOI: http://dx.doi.org/10/gf3kch

[181] Hemant Bhaskar Surale, Aakar Gupta, Mark Hancock, and Daniel Vogel. 2019a. TabletInVR: Exploring the Design Space for Using a Multi-Touch Tablet in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, Glasgow, Scotland Uk, 13:1–13:13. DOI:http://dx.doi.org/10/gf2d42

[182] Hemant Bhaskar Surale, Fabrice Matulic, and Daniel Vogel. 2019b. Experimental Analysis of Barehand Mid-Air Mode-Switching Techniques in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, Glasgow, Scotland Uk, 196:1–196:14. DOI: http://dx.doi.org/10/gf3kx2

[183] Penelope Sweetser and Peta Wyeth. 2005. GameFlow: A Model for Evaluating Player Enjoyment in Games. *Comput. Entertain.* 3, 3 (July 2005), 3–3. DOI: http://dx.doi.org/10/d4k4x4

[184] Theophilus Teo, Louise Lawrence, Gun A. Lee, Mark Billinghurst, and Matt Adcock. 2019. Mixed Reality Remote Collaboration Combining 360 Video and 3D Reconstruction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, Glasgow, Scotland Uk, 201:1–201:14. DOI:http://dx.doi.org/10/gf2bzn

[185] Balasaravanan Thoravi Kumaravel, Cuong Nguyen, Stephen DiVerdi, and Björn Hartmann. 2019. TutoriVR: A Video-Based Tutorial System for Design Applications in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, Glasgow, Scotland Uk, 284:1–284:12. DOI:http://dx.doi.org/10/gf3krs

[186] Tanh Quang Tran, HyunJu Shin, Wolfgang Stuerzlinger, and JungHyun Han. 2017. Effects of Virtual Arm Representations on Interaction in Virtual Environments. ACM Press, 1–9. DOI: http://dx.doi.org/10.1145/3139131.3139149

[187] Sam Tregillus, Majed Al Zayer, and Eelke Folmer. 2017. Handsfree Omnidirectional VR Navigation Using Head Tilt. ACM Press, 4063–4068. DOI: http://dx.doi.org/10.1145/3025453.3025521

[188] Hsin-Ruey Tsai, Jun Rekimoto, and Bing-Yu Chen. 2019. ElasticVR: Providing Multilevel Continuously-Changing Resistive Force and Instant Impact Using Elasticity for VR. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, Glasgow, Scotland Uk, 220:1–220:10. DOI:http://dx.doi.org/10/gf2hjr

[189] Martin Usoh, Ernest Catena, Sima Arman, and Mel Slater. 2000. Using Presence Questionnaires in Reality. *Presence: Teleoperators and Virtual Environments* 9, 5 (2000), 497–503. DOI:http://dx.doi.org/10/bf3cmp

[190] Adrien Verhulst, Jean-Marie Normand, Cindy Lombart, and Guillaume Moreau. 2017. A Study on the Use of an Immersive Virtual Reality Store to Investigate Consumer Perceptions and Purchase Behavior toward Non-Standard Fruits and Vegetables. IEEE, 55–63. DOI:http://dx.doi.org/10.1109/VR.2017.7892231

[191] Alexandra Voit, Sven Mayer, Valentin Schwind, and Niels Henze. 2019. Online, VR, AR, Lab, and In-Situ: Comparison of Research Methods to Evaluate Smart Artifacts. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, Glasgow, Scotland Uk, 507:1–507:12. DOI: http://dx.doi.org/10.1145/3290605.3300737

[192] Emanuel Vonach, Clemens Gatterer, and Hannes Kaufmann. 2017. VRRobot: Robot Actuated Props in an Infinite Virtual Environment. IEEE, 74–83. DOI: http://dx.doi.org/10.1109/VR.2017.7892233

[193] Daniel Wagner, Christian Hofmann, Heiko Hamann, and Sebastian von Mammen. 2017. Design and Exploration of Braiding Swarms in VR. ACM Press, 1–4. DOI:http://dx.doi.org/10.1145/3139131.3139169

[194] Ryan Wedoff, Lindsay Ball, Amelia Wang, Yi Xuan Khoo, Lauren Lieberman, and Kyle Rector. 2019.

Virtual Showdown: An Accessible Virtual Reality Game with Scaffolds for Youth with Visual Impairments. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, Glasgow, Scotland Uk, 141:1–141:15. DOI:http://dx.doi.org/10/gf2nfp

[195] Florian Weidner, Anne Hoesch, Sandra Poeschl, and Wolfgang Broll. 2017. Comparing VR and Non-VR Driving Simulations: An Experimental User Study. IEEE, 281–282. DOI: http://dx.doi.org/10.1109/VR.2017.7892286

[196] T. Weißker, A. Kunert, B. Fröhlich, and A. Kulik. 2018. Spatial Updating and Simulator Sickness During Steering and Jumping in Immersive Virtual Environments. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 97–104. DOI: http://dx.doi.org/10/gf39qj

[197] Martin Westhoven, Dennis Paul, and Thomas Alexander. 2016. Head Turn Scaling below the Threshold of Perception in Immersive Virtual Environments. ACM Press, 77–86. DOI: http://dx.doi.org/10.1145/2993369.2993385

[198] C. Wienrich, K. Schindler, N. Döllinqer, S. Kock, and O. Traupe. 2018. Social Presence and Cooperation in Large-Scale Multi-User Virtual Reality - The Relevance of Social Interdependence for Location-Based Environments. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 207–214. DOI:http://dx.doi.org/10/gf39qb

[199] Graham Wilson and Mark McGill. 2018. Violent Video Games in Virtual Reality: Re-Evaluating the Impact and Rating of Interactive Experiences. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '18)*. ACM, Melbourne, VIC, Australia, 535–548. DOI: http://dx.doi.org/10/gf3ngh

[200] Graham Wilson, Mark McGill, Matthew Jamieson, Julie R. Williamson, and Stephen A. Brewster. 2018. Object Manipulation in Virtual Reality Under Increasing Levels of Translational Gain. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 99:1–99:13. DOI: http://dx.doi.org/10.1145/3173574.3173673

[201] Bob G. Witmer and Michael J. Singer. 1998. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments* 7, 3 (June 1998), 225–240. DOI: http://dx.doi.org/10/dw9b5s

[202] Soojeong Yoo and Judy Kay. 2016. VRun: Running-in-Place Virtual Reality Exergame. In *Proceedings of the 28th Australian Conference on Computer-Human Interaction (OzCHI '16)*. ACM, New York, NY, USA, 562–566. DOI: http://dx.doi.org/10.1145/3010915.3010987

[203] Majed Al Zayer. 2019. *Universal Usability of Virtual Reality*. Dissertation. University of Nevada, Reno, Nevada, Reno.

[204] André Zenner and Antonio Krüger. 2019. Drag:On: A Virtual Reality Controller Providing Haptic Feedback Based on Drag and Weight Shift. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, Glasgow, Scotland Uk, 211:1–211:12. DOI:http://dx.doi.org/10/gf2hjs

[205] Yuhang Zhao, Edward Cutrell, Christian Holz, Meredith Ringel Morris, Eyal Ofek, and Andrew D. Wilson. 2019. SeeingVR: A Set of Tools to Make Virtual Reality More Accessible to People with Low Vision. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, Glasgow, Scotland Uk, 111:1–111:14. DOI: http://dx.doi.org/10/gf2nfg

Publication P4

Challenges of Controlling the Rotation of Virtual Objects with Variable Grip Using Force-Feedback Gloves

Michael Bonfert, Maiko Hübinger, and Rainer Malaka

This publication investigates an interaction technique replicating the variable grip strength applied to a held object using force-feedback gloves in VR. We map the exerted finger pressure to the rotational freedom of the virtual object for a firm or loose grip. A user study (N=21) showed how challenging it was for participants to control the object's rotation with our prototype. The grip variability led to poorer performance and increased task load compared to the default fixed rotation. We discuss system limitations and how to overcome them in future haptic interfaces.

My contribution: I contributed the research idea (*conceptualization*) and wrote the first draft of the manuscript (*writing – draft*). I contributed equally to the design of the interaction technique (*conceptualization*), study design (*methodology*), statistical analysis (*formal analysis*), the revision of the manuscript (*writing – review & editing*), and the figures (*visualization*) as MH. I made minor contributions to the prototype development (*software*). I supervised the project (*supervision*).

Published in *Frontiers in Virtual Reality*, Section *VR and Human Behaviour*, Research Topic *Beyond Touch: Free Hand Interaction in Virtual Environments*, 2023 frontiers | Frontiers in Virtual Reality

TYPE Original Research
PUBLISHED 09 October 2023
DOI 10.3389/frvir.2023.1190426

OPEN ACCESS

EDITED BY
Florian Daiber,
German Research Center for Artificial Intelligence (DFKI), Germany

REVIEWED BY
André Zenner,
Saarland University, Germany
Sebastian Friston,
University College London,
United Kingdom

*CORRESPONDENCE
Michael Bonfert,
bonfert@uni-bremen.de

RECEIVED 20 March 2023
ACCEPTED 08 September 2023
PUBLISHED 09 October 2023

CITATION
Bonfert M, Hübinger M and Malaka R (2023), Challenges of controlling the rotation of virtual objects with variable grip using force-feedback gloves.
Front. Virtual Real. 4:1190426.
doi: 10.3389/frvir.2023.1190426

Challenges of controlling the rotation of virtual objects with variable grip using force-feedback gloves

Michael Bonfert[1]*, Maiko Hübinger[2] and Rainer Malaka[1]

[1]Digital Media Lab, University of Bremen, Bremen, Germany, [2]University of Bremen, Bremen, Germany

Some virtual reality (VR) applications require true-to-life object manipulation, such as for training or teleoperation. We investigate an interaction technique that replicates the variable grip strength applied to a held object when using force-feedback gloves in VR. We map the exerted finger pressure to the rotational freedom of the virtual object. With a firm grip, the object's orientation is fixed to the hand. With a loose grip, the user can allow the object to rotate freely within the hand. A user study (N = 21) showed how challenging it was for participants to control the object's rotation with our prototype employing the SenseGlove DK1. Despite high action fidelity, the grip variability led to poorer performance and increased task load compared to the default fixed rotation. We suspect low haptic fidelity as an explanation as only kinesthetic forces but no cutaneous cues are rendered. We discuss the system design limitations and how to overcome them in future haptic interfaces for physics-based multi-finger object manipulation.

KEYWORDS

object manipulation, haptics, force feedback, virtual reality, XR, dexterity

1 Introduction

When we encounter objects in virtual reality (VR), we intuitively attempt to handle the virtual objects as if they were real. We must soon realize that our manual abilities in VR are comparatively limited. However, it is essential in some VR applications to reproduce reality with high fidelity. For example, the simulation of interactions must be as realistic as possible for training motor skills, such as in medicine, manufacturing, or engineering, but also for robot and surgery teleoperation, digital twins, digital marketing, or mixed reality. These use cases require a precise one-to-one mapping of the users' actions to the virtual hand enabling true-to-life object manipulation. To simulate the astonishing dexterity of the human hands, the control over an object must go beyond having it glued to the virtual hand as a single-point effector. With current VR systems, when we hold a virtual object and move or turn our hand, the movement and rotation are often transferred directly to the object. The object's possible orientations are thus limited by how far we can revolve our hand. Beyond this extrinsic movement of turning the entire hand, we would control a real object with intrinsic movements, i.e., use our fingers to adjust how we hold it within our hand (Elliott and Connolly, 1984).

We can impact a held object through the pressure we apply to it. With our grip, we determine how our skin and the object's surface interact. For example, when holding an object with only gentle pressure, we can let it slip through our fingers. This has been virtually recreated with a haptic device that renders the tangential forces of the sliding object to the user's fingertips (Kim et al., 2022). Previous work also explored the possibilities of applying pressure as means of input.

Using squeezing actions on haptic devices has been investigated in abstract ways, e.g., when holding a smartphone (Yabe et al., 2017; Quinn et al., 2019), and with pinch gestures in mixed reality applications (Schmitz et al., 2022), but also for basic object manipulation (Achibet et al., 2014). Previous work also addressed the experience of object elasticity in VR by exerting pressure on the deformable object with controllers (Tsai et al., 2019), via a proxy (Ryu et al., 2020), or with force-feedback gloves (Coquillart et al., 2004).

In this work, we expand on the interaction technique proposed by Bonfert et al. (2019), which considers the grip applied to a held object using controller-based systems. By varying the grip strength, the object can be held loosely, allowing it to rotate freely between the fingers, or it can be grasped firmly, which transfers the hand's rotation directly to the object. This grip variability affords control over additional rotational degrees of freedom. Without it, users must release and grasp the object again in the desired orientation, called clutching (Zhai et al., 1996). The affordances of grip variability influence our everyday tasks in various situations—often subconsciously—e.g., when moving a full glass, using a screwdriver, screwing in a light bulb, or fidgeting with a pen. Imagine moving a glass of water from a higher shelf to a table without gravity, ensuring the opening is always level to the top. The control over the grip strength allows a flexible grasp with automatic vertical alignment so that nothing is spilled. In other situations, we rely on an object's inertia to change its rotation. For example, we would intuitively swing a book in an upright orientation by leveraging its momentum with a loose grip before putting it on a shelf. Virtually reproducing these natural grip-based hand manipulations could benefit use cases that require high interaction fidelity. Beyond realistic interactions, a variable grip can compensate for inadequate object orientations due to the initial grasp, e.g., to finely adjust the angle between the hand and a virtual hammer or knife.

An evaluation of the interaction technique showed advantages in terms of user satisfaction, intuitiveness, and realism with a slight decrease in the task load (Bonfert et al., 2019). This study used a controller-based system with HTC Vive controllers for input. Users set the grip strength with the trigger or grip buttons depending on the experiment condition. The abstract button actions must be mapped to the intended virtual actions of controlling an object, which was found to increase mental demand (Bonfert et al., 2019).

The interaction technique was implemented for the Valve Index controllers in a follow-up study by Pedersen et al. (2023). The user can grab virtual objects with these controllers by enclosing and pressing the controller's handle. The pressure applied to the handle was interpreted as direct input for the grip strength applied to the object. This removed the need for mapping, which increases input fidelity. In line with previous findings (Bonfert et al., 2019), an evaluation showed that the interactions with variable grip were considered the most realistic, slightly slower, and with no difference in the TLX ratings. Only the higher usability could not be confirmed in this study (Pedersen et al., 2023).

1.1 Finger-based manipulation with force feedback

For applying pressure to something virtual, rendering a resistance force from the object's surface is helpful. There is a rich history of research on haptic devices providing kinesthetic forces to the user's hand and fingers (Massie and Salisbury, 1994; Hirota and Hirose, 1995; Bouzit et al., 2002; Minamizawa et al., 2007; Endo et al., 2011), typically for translational manipulations, shape exploration, or weight simulation. A recent survey on glove-shaped haptic devices that render force feedback details different designs and their characteristics (Wang et al., 2019).

VR systems with force-feedback gloves provide precise hand poses with input actions from the individual fingers. This direct form of grasping exactly resembles the hand movements of handling real objects. Additionally, the user is provided with haptic feedback that simulates the object's resistance when pressing against its surface. The user can vary the finger pressure to adjust the grip strength as illustrated in Figure 1. Thus, the proposed interaction technique combines isomorphic control over object selection and movement with isometric control over rotational freedom. The higher input and feedback fidelity of using force-feedback gloves might enable a more natural control over a held object. Therefore, we developed a system that enables the user to control the rotation of virtual objects with variable grip through a dorsal-based SenseGlove DK1 with force feedback, as shown in Figure 1. In a quantitative user study ($N = 21$), we evaluated the users' experiences with the system regarding placement accuracy, the time needed, the number of grasps, self-reports on task load, perceived object control, and aspects of presence. In a pick-and-place task, users moved 36 objects to a target area with a given position and orientation.

Although we anticipated higher perceived realism and user satisfaction with similar performance, we found disadvantages of providing grip variability. The participants performed poorly and reported a high task load even in the baseline condition. We discuss the experiment outcome and provide possible explanations why the interaction technique fell short of expectations. We derive opportunities for future research that builds on our experiences. Overall, we contribute a system design that advances the interaction technique of grip variability to glove-based interaction, as well as the results and lessons learned from an initial prototype evaluation for finger pressure-based object manipulations in VR.

2 System design

We built an interaction system for our user study using the SenseGlove DK1 VR gloves (SenseGlove, 2023). This exoskeleton is grounded on the back of the hand and connected with Velcro straps to the fingertips. It weighs 300 g per glove. The DK1 can restrict each finger's inward movement to simulate the contact force of an object's surface. It can only inhibit finger movement but not alter its position. Cables run from the glove's base up to each finger segment's tip and through the segment's joints. Applying a brake force on the cable prevents the finger from moving further inwards (flexing). The hand position is captured with an HTC Vive Tracker. The glove precisely tracks the fingers' pose and movements. In order to create the 3D representation of the glove, the angles of the four linkages of each segment are measured at a rate of 120 Hz and a resolution of 0.35°. Based on this, the SenseGlove software calculates the brakes' resistances. For each finger, a maximum force of 40 N can be applied at the fingertip which can be updated at a rate of up to 200 Hz and a resolution of 100 steps of force (SenseGlove, 2019). The resistance increases when a collider on the fingertip approaches

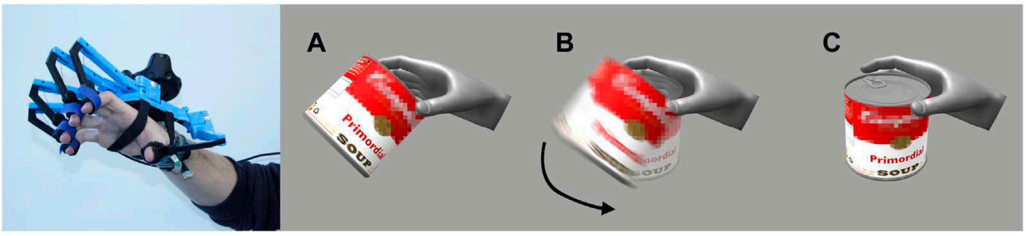


FIGURE 1
A user changes the orientation of a virtual soup can. Left: The SenseGlove DK1 is attached to the user's fingers providing force feedback along the can's shape. Right: **(A)** The user applies a firm grip. Thus the rotation of the can is fixed. **(B)** When the user loosens the grip by reducing the finger pressure, the can swings downwards due to gravity. **(C)** The can stays in a level orientation. Created with Unity Editor®. Unity is a trademark or registered trademark of Unity Technologies.

and interpenetrates an item. A maximum penetration depth and a force value for the maximum depth can be configured for the item. Based on that, a force value is calculated depending on how deep the collider is interpenetrating the item.

For the experiment, we used a Valve Index headset. The virtual scene was a hobby workshop with a workbench. The research question is on interactions with high simulation fidelity in terms of physical realism, so we aimed for photo-realistic objects. It was built in Unity 2020.3.23f1 using the SteamVR Unity Plugin v2.7.3 (SDK 1.14.15) and the SenseGlove Unity Plugin v2.3.1. We modified the plugin's grasping interaction for the different experimental conditions. When the user reaches for an object, it is firmly attached to the hand as soon as enough fingers touch the surface and press against it. As the SenseGlove interaction system necessarily attaches the held object in fixed relation to the hand, we spawn a copy of the object at grasp. The copy is fully visible and seems to the user like the object that is actually being held and manipulated. Directly after grasping, it is aligned with the original object and will always follow its position. But when using a loose grip, the copy's orientation deviates from the original orientation due to gravity. It can rotate in any direction and neither collides with the original object nor the hand, thus potentially penetrating the hand model. The potential hand penetrations reduce simulation fidelity, but this design decision allowed more rotational freedom increasing the interaction technique's utility. Depending on how the object was grasped initially, the behavior is also better predictable when enabling interpenetration. As the object's copy is visible to the user, the original instance of the object is invisible while it is grasped. The original shape still provides force feedback to the fingers. Hence, the resistance forces do not fluctuate while the object's copy is rotating, as the held object always remains in its original orientation relative to the hand.

For controlling the object's freedom of rotation, we interpreted the glove's resistance force as pressure against the object's surface. The softer the grip, the more freely the object can rotate within the hand. A loose object rotates with three degrees of freedom around an anchor point in the middle between the fingertips of the thumb and index finger, marked with an x in Figure 2. While an axis as the center of rotation would be more realistic in most cases, given the grasp with an opposing thumb, it would also restrict the versatility and predictability of the interaction technique. Therefore, the object's rotation was not constrained in any direction when grasped softly. More pressure restricted the rotational freedom until fully locked to the hand's rotation using a firm grip.

We implemented this mechanism in Unity with a *Configurable Joint* attached to the held object and connected to the copy of the object, which feels rather rusty or slack depending on the *Slerp Drive–Position Spring* parameter. The lower the position spring value is, the less the copy object resists gravity; hence, the more it can deviate from the original object's orientation. This is visualized in Figure 2. The object's elasticity was not considered for the calculations. Consequently, flexible objects (such as the milk carton in the user study) did not soften or yield when grasped firmly.

After pilot testing, we found that the perceived control over the grip input was not as precise as hoped. When trying to grip it as gently as possible, the object was at risk of falling. Therefore, we simplified the interaction design to binary grip states: above 80% of finger pressure, the rotation was locked; below that, rotation is possible with slight resistance. This was visualized to the user with a bar above the hand that was filled with higher pressure. Beyond the threshold, it changed from a green to a red area indicating the firm grip, as shown in Figure 3. It was visible when holding an object in the condition with variable grip. Further, for varying the grip strength, we initially considered the pressure input of all fingers. After this turned out unreliable in informal testing, we only used the thumb and index finger pressure to control the grip. Nevertheless, all fingers still received force feedback.

3 Study design

We evaluated our proposed interaction technique in a user study with 21 participants. The pick-and-place tasks in the experiment are inspired by the study by Bonfert et al. (2019), but the study design has been adjusted. The experiment had a within-subject design comparing the two conditions *fixed grip* and *variable grip*. In the condition with a fixed grip, the users experienced a standard interaction design of attaching an object firmly to the hand when grasping it. The object's rotation is directly linked to the rotation of the hand. To change the angle between the object and the hand, the user must release the object and grab it again. In the condition with variable grip, the user experiences our new interaction technique that allows dynamic control over the object's rotation. Adjusting the strength with which the fingers press against the virtual object either fixes the rotation as in the other condition or releases the object to rotate freely. With a loose grip, the object's center of mass would

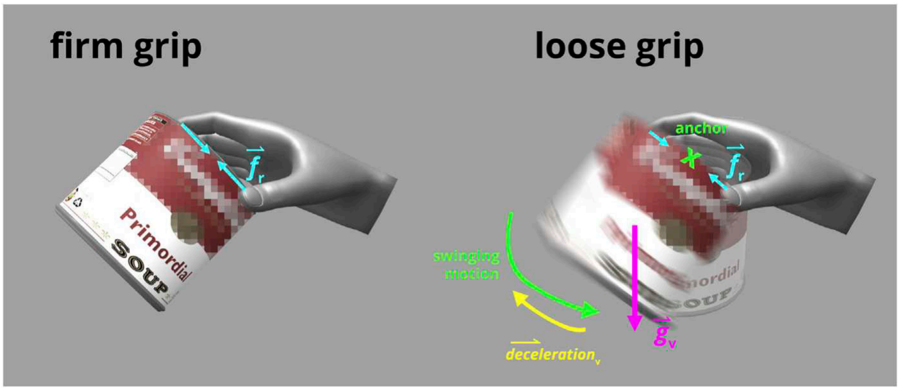


FIGURE 2
Forces and constraints during the interaction: forces in reality $\vec{f_r}$ from the fingers pressing against the glove (cyan), virtual gravitational force $\vec{g_v}$ (magenta), the *swinging motion* of the object due to gravity (green), virtual anchor point x from the *configurable joint* around which the object rotates (green), and the virtual counter force $\overrightarrow{deceleration_v}$ from the *slerp drive* component slowing down the swinging motion. Created with Unity Editor®. Unity is a trademark or registered trademark of Unity Technologies.

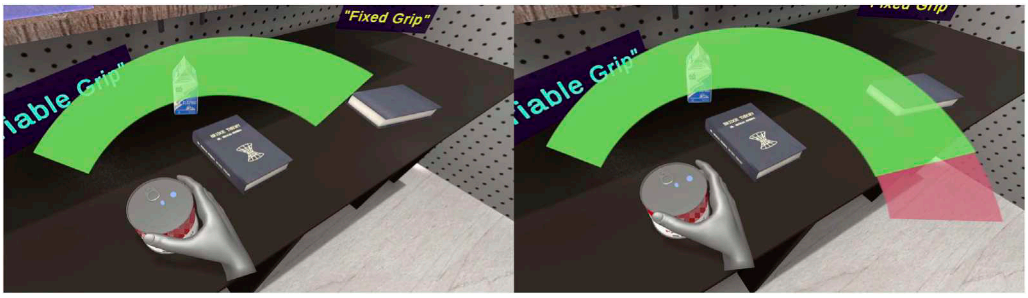


FIGURE 3
The indicator above the user's hand visualizes the grip strength applied to the object. The user grasps the object firmly when the bar is filled to the red area. Created with Unity Editor®. Unity is a trademark or registered trademark of Unity Technologies.

FIGURE 4
The three object types used in task B of the experiment: a can (diameter of 11 cm, 10 cm high), a book (20 cm × 14 cm x 4 cm), and a milk carton (base of 7 cm × 8 cm, 16 cm high). Created with Unity Editor®. Unity is a trademark or registered trademark of Unity Technologies.

rotate downwards following the gravitational pull, independent of the hand's rotational movement. All participants tested both conditions in counterbalanced order. Although including two additional conditions with the controller-based implementations from Bonfert et al. (2019) would have been interesting for comparison, we refrained from prolonging the study further due to the already required time, object manipulations, and filled questionnaires.

3.1 Tasks

The participants were asked to pick up various objects and move them to a target in two tasks per condition. The objects needed to be placed in the correct location, with a deviation of less than 3 cm, and with the correct orientation, deviating less than 20°. In task A, six cans had to be moved to a target. The poses (position and orientation) of both the start and the target were identical for all six cans. This resulted in repeated measures with similar trajectories. After a can was correctly placed within the thresholds, it would disappear after 0.5 s, and the next one would appear. In task B, three types of objects were moved: cans, books, and milk cartons, as shown in Figure 4. Four instances of every object type had to be moved from different starting poses to the identical target pose, as illustrated in Figure 5. The wide range of starting poses required the users to vary trajectories and rotations between the objects. While the starting

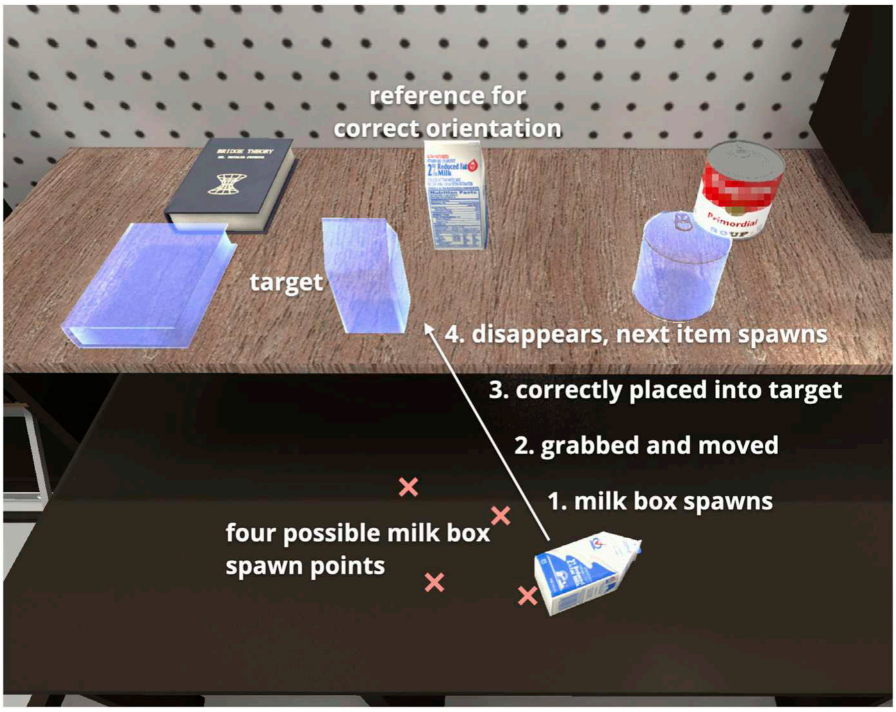


FIGURE 5
The steps of moving an item to its target in task B. There is one item to be moved at a time. Each of the twelve items will appear in an individual orientation and point, although these poses are predefined and equal between the different trials. A reference that does not physically interfere is placed behind the target area so participants can derive the desired orientation. Created with Unity Editor®. Unity is a trademark or registered trademark of Unity Technologies.

poses were identical for all participants, the order of the objects was randomized. However, each participant had the same order for both conditions. The order of the tasks was counterbalanced. In total, 756 object manipulations were measured ((6 Task-A objects+12 Task-B objects)*2 conditions * 21 participants). After cleaning the data from outliers (3*IQR) and system errors, 654 valid cases could be used for analysis. Due to the repeated-measures design, 287 pairs yield valid data in both conditions.

3.2 Procedure

After giving informed consent and filling in a demographic questionnaire, the participants were introduced to the VR equipment, including the headset and force-feedback glove. Only one glove was used to prevent bi-manual interactions. For increased comparability, all participants used the right glove and were right-handed. They entered a virtual workshop environment with tables and shelves where the tutorial and experiment took place. Tooltips with instructions and explanations guided them through the process. Before each condition, the participants could test the interaction mode in a tutorial until they felt confident and demonstrated proficiency in both interaction modes with all three items. Then, they performed both tasks and filled in the questionnaires, followed by the same process in the other condition. The questionnaires included the *Presence Questionnaire* by Witmer and Singer (1998), the raw *NASA TLX* by Hart and Staveland (1988), and five custom questions. The following custom items were rated on a scale of 1–7, with the labels in brackets.

1. Compared to the real world, how PRECISELY could you place objects in the demanded location and orientation? (*much less precisely* to *much more precisely*)
2. Compared to the real world, how FAST could you place objects in the demanded location and orientation? (*much slower* to *much faster*)
3. I was aware of how tight I was gripping the items. (*strongly disagree* to *strongly agree*)
4. I could develop a sense of how tight I was gripping the items. (*strongly disagree* to *strongly agree*)
5. I could move and rotate the items as I expected. (*strongly disagree* to *strongly agree*)

They were integrated into the virtual environment and operated with an HTC Vive controller given to the participants in their free hand. Following the recommendations by Alexandrovsky et al. (2020), the in-VR questionnaires avoided interrupting the VR experience and might improve data quality. After the measurements, the participants could return to the tutorial and keep testing the technology. In the end, a semi-structured interview over approximately 5 minutes was conducted. The experiment, questionnaires, and interviews were held in English. If preferred by the participant, the interview was held in German. In total, the experiment took approximately 45 min on average.

3.3 Sample

The sample was recruited on the university campus with email, leaflets, and word-of-mouth advertisements. There was no

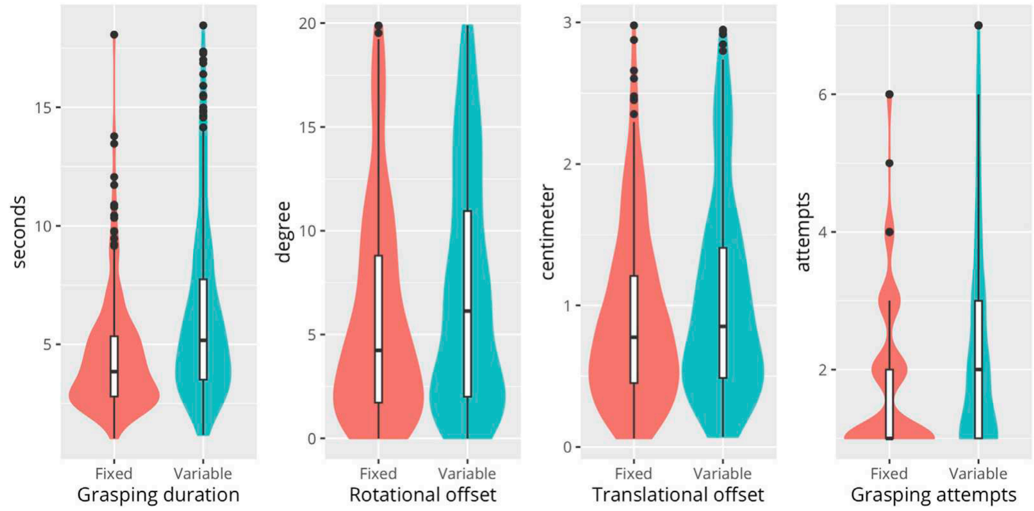


FIGURE 6
Violin plots for the performance data on participants' average grasping duration, rotational offset, translational offset, and grasping attempts per condition.

financial compensation for participating. 21 people participated in the experiment, of which six self-identified as female and 15 as male. Their age ranged from 15 to 61, averaging 27.3 years. All participants were right-handed and used the right glove for the experiment. Nine participants had never used VR before the experiment, and only two used VR at least every month. Four people reported feeling moderately or very experienced with VR, seven little, and ten not at all. Ten participants had used VR controllers before, of which eight used them for object manipulation. Only four have already used some form of hand tracking, such as gloves or optical finger tracking, of which three used it for handling objects. Thus, it was the first time for 18 participants to use a glove to move around objects.

3.4 Data analysis

All statistical tests are calculated with an alpha level of .05, Bonferroni-Holm-corrected, and two-sided assuming any difference between the conditions. The distribution of the metric data from the performance measurements deviates from normality according to the Shapiro-Wilk-Test (p = [0.001 .. 0.061]). Therefore, we used the non-parametric Wilcoxon signed-rank test to find group differences. It was also used for the ordinal data collected in the questionnaires. The effect sizes are reported as matched pairs rank biserial (r_{rb}) and can be interpreted as a correlation coefficient. We checked the internal consistency with reliability analysis because the custom questionnaire items are not validated as standardized scales. The tests yielded Cronbach's Alpha values of $\alpha = .71$, indicating that the single items describe the same underlying concept. The cleaned quantitative data with test reports are available in a repository on OSF (see section 6). The qualitative data from the semi-structured interviews were analyzed unsystematically to identify relevant insights that might explain the quantitative findings.

4 Results

Overall, the participants performed better in the *fixed grip* condition. The distributions of the performance data are shown in Figure 6. The translational accuracy of placing the objects on the targets was higher with a fixed grip ($Median_f = 8\ mm$ offset $\pm\ SD_f = 6\ mm$) than with a variable grip ($Mdn_v = 9 \pm 7\ mm$). Hence, the distance from an object to the target's center was 1 mm smaller, which is statistically significant with a small effect size ($Z = -2.4$, $p = .015$, $r_{rb} = -.17$). The rotational accuracy was also higher with a fixed grip ($Mdn_f = 4.2^\circ \pm 5^\circ$ deviation, $Mdn_v = 6.1^\circ \pm 5.6^\circ$). The difference of 1.9° in orientation accuracy is significant with a small effect size ($Z = -3.2$, $p < .005$, $r_{rb} = -.22$). Further, the participants needed less **time** to successfully place an object when using a fixed grip ($Mdn_f = 3.8 \pm 2.3\ s$, $Mdn_v = 5.2 \pm 3.6\ s$). This 1.4 s difference is significant with a large effect size ($Z = -3.4$, $p < .005$, $r_{rb} = -.57$). Similarly, the participants needed fewer **attempts** with a fixed grip ($Mdn_f = 1 \pm 1.1$ grasps per object, $Mdn_v = 2 \pm 1.5$ grasps), which is a significant difference with a medium effect size ($Z = -4.6$, $p < .005$, $r_{rb} = -.43$).

The interaction technique with a fixed grip was generally rated better in the questionnaires. The overall score of the Presence Questionnaire (Witmer and Singer, 1998) and also its subscores showed no group differences ($p > .805$), except for the subscale *interface quality*. Here, the participants rated how much the interface interfered with the task with a significant advantage of using a fixed grip ($Mdn_f = 12 \pm 2.4$, $Mdn_v = 10 \pm 2$) showing a large effect size ($Z = 2.9$, $p = .018$, $r_{rb} = .77$). While the raw score of the NASA TLX (Hart and Staveland, 1988) indicates a lower workload of using the fixed grip ($Mdn_f = 41.7 \pm 9$, $Mdn_v = 43.3 \pm 6.4$), this difference is not significant after Bonferroni-Holm correction ($p = .092$). However, there are significant differences with large effect sizes for the items *mental demand* ($Z = -3.1$, $p = .014$, $r_{rb} = -.89$), *performance* ($Z = -3.0$, $p = .018$, $r_{rb} = -.76$), and *frustration* ($Z = -2.9$, $p = .02$, $r_{rb} = -.78$), each with higher demands for variable grip. The distributions of the most insightful TLX items are shown in

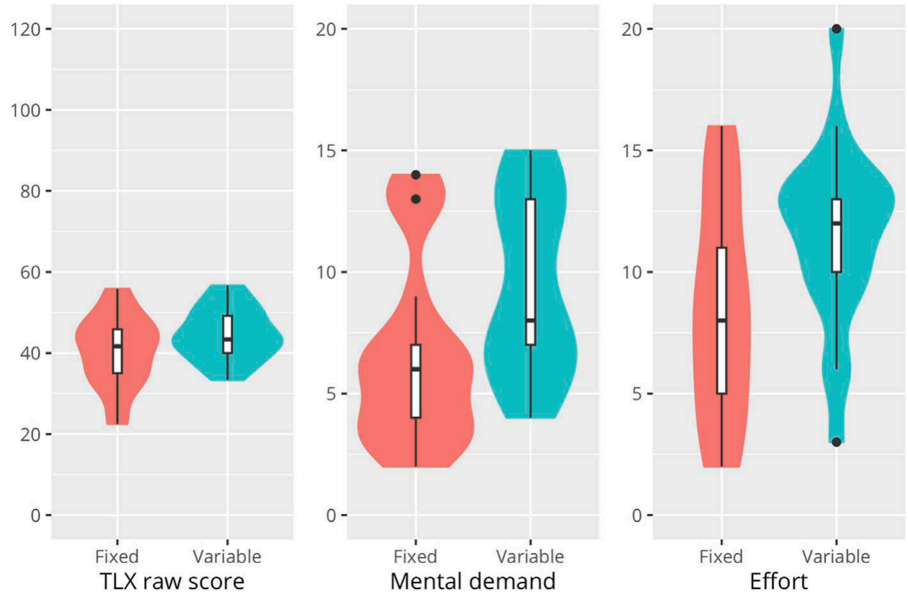


FIGURE 7
Selected violin plots for the NASA Task Load Index raw score and the items *mental demand* and *effort* per condition.

Figure 7. Only one of the custom questionnaire items was rated significantly differently. Participants gave a higher rating on whether they "could move and rotate the items" as they expected ($Mdn_f = 6 \pm 1.2$, $Mdn_v = 5 \pm 1.5$) with a large effect size ($Z = 2.9$, $p = .018$, $r_{rb} = .79$).

5 Discussion

The study results showed a poorer performance with the grip variability in a virtual pick-and-place task using a force-feedback glove. The object placement was slightly less accurate, but the study participants needed 40% longer and considerably more grasping attempts to move an object. This is in line with previous research on controller-based systems in which varying the grip of a held object was found to take more time and sometimes additional grabs (Bonfert et al., 2019; Pedersen et al., 2023). In contrast to the outcome of this experiment, grip variability with controllers was reported to be more intuitive, easier to control, and more satisfactory (Bonfert et al., 2019), as well as more realistic (Bonfert et al., 2019; Pedersen et al., 2023), compared to the condition with a fixed grip. However, when using a force-feedback glove with a variable grip, participants reported higher mental load and frustration, inferior perceived performance, and less agency over the object behavior.

There are several possible explanations for these unexpected results. One reason for the poor performance and ratings of the proposed interaction technique might be the general difficulties in handling objects with the SenseGlove DK1 as an early development prototype. Considering the simple task, we measured relatively high TLX scores and slow placement even in the baseline condition with a simple pick-and-place task. According to interview feedback, many users did not feel confident using the glove.

Its force feedback mechanism posed a severe limitation in the experiment as the rendered forces depend on the angle at which a finger presses against the surface. A perfectly perpendicular angle results in appropriate resistance, but the more oblique the finger touches the surface, the higher the chance of no or unexpected feedback. While this is already confusing with the fixed-grip baseline, it is even more confusing with a variable grip. A change in pressure against the surface controls the change between a firm and a loose grip. If the resistance from the surface is unexpected and inadequate, it is impossible to utilize it for intuitive control over the grip strength. As another mechanical restriction of the SenseGlove DK1, the applied pressure can only be approximated from the glove's generated resistance. Readings from pressure sensors at the fingertips might provide more accurate data and allow finer control over the grip strength.

Another reason might be our sample with many novice users. 18 out of 21 participants have never used hand tracking for object manipulation before. Already challenged by the glove's behavior, they mostly restricted their actions to the simpler fixed grip, which resulted in a more predictable outcome, even if it required more manual movement and inconvenient hand poses. Similar to an interaction technique with continuous variable grip tested by Bonfert et al. (2019), in which the grip strength was set with a controller's trigger button, the grip adjustment with the glove was too delicate. Due to the steep learning curve, participants assumed they could have performed better with considerably more practice and might then find it helpful.

From our observations of the experiments and statements in the interviews, we suspect one more reason for the slower object handling times of the variable grip. The unsteady grasp of the variable grip mode caused more items to be accidentally dropped by the participants, which naturally resulted in longer handling times on average because the items had to be grabbed again.

Nonetheless, once the users established a secure grasp, moving the item along the required trajectory and placing it down was not observed to be slower with the variable grip.

An additional limitation of our implementation was the visual indicator of the currently applied grip strength. Participants described it as rather distracting than helpful. Alternative cues might be considered in future work, but ideally, the only visual indicator necessary is the observed object behavior. Ideally, users should be able to feel the object's state only from haptic feedback in a system reproducing grip-based manipulation of real objects with high interaction fidelity.

An improved physics simulation with the reliable rendering of kinesthetic forces might enable more confident control over holding the object. However, even when rendered perfectly, it might be insufficient for users to intuitively control the applied grip strength. Surface-finger interaction involves other physical forces imperative for high haptic fidelity (Muender et al., 2022). When handling real objects, humans are excellent at maintaining a balance between a grip strong enough to prevent slipping yet not excessively powerful (Westling and Johansson, 1984). However, to achieve this, we interpret the frictional sensation at the fingertips (Cadoret and Smith, 1996), which our prototype does not render. When gripping an object, the frictional condition is informed by tangential, not kinesthetic forces (Augurelle et al., 2003) such as skin stretches. If cutaneous cues are missing, as studies have shown using local anesthesia, people also drop real objects or use an overly powerful grip because their mental model of the held object's physical properties is insufficiently informed (Westling and Johansson, 1984; Augurelle et al., 2003).

Overall, the presented study has not shown that mapping the finger pressure with force-feedback gloves to the grip strength of handling virtual objects would, in principle, be an undesirable solution. We have not yet demonstrated the potential we anticipate when using hardware with more accurate actuation, sophisticated force vector estimation of individual fingers, and additional cutaneous feedback.

5.1 Future work

Therefore, supporting tactile cues beyond kinesthetic forces must be considered for multi-finger object manipulation when controlling grip strength. Shear forces, friction, slip, or contact forces have been shown to improve reproducing real-world haptic experiences (Girard et al., 2016; Whitmire et al., 2018; Salazar et al., 2020; Kim et al., 2022). Also, the impact of their combination with kinesthetic feedback from gloves would be insightful. The ideal prototype for exploring the benefits of natural grip-based object manipulation would render detailed tactile cues, rely on complex physics calculations comprising the individual pressure of all fingers, afford dynamic and continuous grip, and avoid interpenetration of the object and the hand.

While in this study, we explored using the interaction technique for objects with physically realistic behavior, it could also be applied with no gravitational forces in use cases that demand flexible control over the rotational degrees of freedom without clutching. By deactivating gravity for the held object, its orientation can be finely adjusted in any direction without the vertical pull of gravity. This could be useful in educational training, e.g., for surgery or assembly tasks. In addition to the performance indicators, user satisfaction and comfort of hand poses could be operationalized more informatively.

Aside from interactions with force-feedback gloves, improvements in optical hand tracking and rich mixed-reality applications offer intriguing opportunities for similar interaction techniques. For this, there is a need for complex physics-based simulations that infer object behavior from the position and properties of the individual fingers (Höll et al., 2018). As the contact points between fingers and surface change according to how the object moves within the hand, the adaption of the grasping pose must be determined dynamically, as previous work explored for freehand grasping (Dalia Blaga et al., 2021).

6 Conclusion

We present a system that allows the user to control the rotation of a held virtual object by adjusting the pressure of the fingers on the object's surface, hence, varying the strength of the grip. The prototype was realized with the dorsal-based SenseGlove DK1 providing force feedback to the individual fingers. Thus, in contrast to previous work, the system directly maps how strongly the user presses against the glove's resistance to how firmly the object is being held. Although we expected a more intuitive and efficient execution of a pick-and-place task, the evaluation shows that users need more time and more attempts, experience a higher task load, and perceive less agency over the object. Technical limitations and theoretical considerations provide explanations and outline possible next steps in research on the dexterous multi-finger manipulation of virtual objects.

Data availability statement

The quantitative dataset with the statistical analysis can be found in the following OSF repository: https://osf.io/d64va.

Ethics statement

Ethical approval was not required for the studies involving humans because the local ethics board only issues approval if funding agencies demand it. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

MB conceived the research idea, wrote the first draft of the manuscript, contributed to the figures, provided feedback on the prototype, and supervised the project. MH implemented the system, conducted the user study, and contributed to the figures and manuscript writing. MB and MH designed the interaction technique, developed the study design, conducted the literature review, and performed the data analysis. RM contributed to the

discussion of the research idea and the manuscript review. All authors contributed to the article and approved the submitted version.

Funding

This research was partially funded by Klaus Tschira Stiftung by financing the scholarship of MB, and the German Federal Ministry of Education and Research (BMBF) as part of the project "Hydrogen for Bremen's Industrial Transformation." Staats- und Universitätsbibliothek Bremen covered the open-access publication fees.

Acknowledgments

We are grateful to Steeven Villa and the Human-Centered Ubiquitous Media group at the Ludwig-Maximilians-Universität München for providing us with the gloves for this research. We thank Tanja Döring and Thomas Münder for their feedback on the prototype and study design.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Achibet, M., Marchal, M., Argelaguet, F., and Lécuyer, A. (2014). "The virtual mitten: A novel interaction paradigm for visuo-haptic manipulation of objects using grip force," in 2014 IEEE Symposium on 3D User Interfaces (3DUI), Minneapolis, MN, USA, 29-30 March 2014 (IEEE), 59–66. doi:10.1109/3DUI.2014.6798843

Alexandrovsky, D., Putze, S., Bonfert, M., Höffner, S., Michelmann, P., Wenig, D., et al. (2020). "Examining design choices of questionnaires in vr user studies," in *Proceedings of the 2020 CHI conference on human factors in computing systems* (New York, NY, USA: Association for Computing Machinery), 1–21. doi:10.1145/3313831.3376260

Augurelle, A. S., Smith, A. M., Lejeune, T., and Thonnard, J. L. (2003). Importance of cutaneous feedback in maintaining a secure grip during manipulation of hand-held objects. *hand-held objects* 89, 665–671. doi:10.1152/jn.00249.2002

Bonfert, M., Porzel, R., and Malaka, R. (2019). "Get a grip! Introducing variable grip for controller-based VR systems," in 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Osaka, Japan, 23-27 March 2019 (IEEE), 604–612. doi:10.1109/VR.2019.8797824

Bouzit, M., Popescu, G., Burdea, G., and Boian, R. (2002). "The rutgers master II-nd force feedback glove," in *Haptic interfaces for virtual environment and teleoperator systems* (Los Alamitos, CA: IEEE Computer Society Press), 145–152. doi:10.1109/HAPTIC.2002.998952

Cadoret, G., and Smith, A. M. (1996). Friction, not texture, dictates grip forces used during object manipulation. *dictates grip forces used Dur. object Manip.* 75, 1963–1969. doi:10.1152/jn.1996.75.5.1963

Coquillart, S., Göbel, M., Maciel, A., Sarni, S., Buchwalder, O., Boulic, R., et al. (2004). "Multi-finger haptic rendering of deformable objects," in *Proceedings of the tenth eurographics conference on virtual environments* (New York, NY, USA: Association for Computing Machinery). doi:10.2312/EGVE/EGVE04/105-112

Dalia Blaga, A., Frutos-Pascual, M., Creed, C., and Williams, I. (2021). "Freehand grasping: an analysis of grasping for docking tasks in virtual reality," in 2021 IEEE Virtual Reality and 3D User Interfaces (VR), Lisboa, Portugal, 27 March 2021 - 01 April 2021 (IEEE), 749–758. doi:10.1109/VR50410.2021.00102

Elliott, J. M., and Connolly, K. J. (1984). A classification of manipulative hand movements. *Dev. Med. Child Neurology* 26, 283–296. doi:10.1111/j.1469-8749.1984.tb04445.x

Endo, T., Kawasaki, H., Mouri, T., Ishigure, Y., Shimomura, H., Matsumura, M., et al. (2011). Five-fingered haptic interface robot: hiro III. *IEEE Trans. haptics* 4, 14–27. doi:10.1109/TOH.2010.62

Girard, A., Marchal, M., Gosselin, F., Chabrier, A., Louveau, F., and Lécuyer, A. (2016). HapTip: displaying haptic shear forces at the fingertips for multi-finger interaction in virtual environments. *Front. ICT* 3, 6. doi:10.3389/fict.2016.00006

Hart, S. G., and Staveland, L. E. (1988). Development of NASA-TLX (task load index): results of empirical and theoretical research. *Adv. Psychol.* 52, 139–183. doi:10.1016/S0166-4115(08)62386-9

Hirota, K., and Hirose, M. (1995). Providing force feedback in virtual environments. *IEEE Comput. Graph. Appl.* 15, 22–30. doi:10.1109/38.403824

Höll, M., Oberweger, M., Arth, C., and Lepetit, V. (2018). "Efficient physics-based implementation for realistic hand-object interaction in virtual reality," in 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Tuebingen/Reutlingen, Germany, 18-22 March 2018 (IEEE), 175–182. doi:10.1109/VR.2018.8448284

Kim, M. J., Ryu, N., Chang, W., Pahud, M., Sinclair, M., and Bianchi, A. (2022). "SpinOcchio: understanding haptic-visual congruency of skin-slip in VR with a dynamic grip controller," in *CHI conference on human factors in computing systems* (New York, NY, USA: Association for Computing Machinery). doi:10.1145/3491102.3517724

Massie, T. H., and Salisbury, J. K. (1994). The phantom haptic interface: A device for probing virtual objects. *Haptic Interfaces Virtual Environ. Teleoperator Syst.* 55 (1), 295–300.

Minamizawa, K., Fukamachi, S., Kajimoto, H., Kawakami, N., and Tachi, S. (2007). "Gravity grabber: wearable haptic display to present virtual mass sensation," in *ACM SIGGRAPH 2007 emerging technologies* (New York, NY, USA: Association for Computing Machinery), 8. doi:10.1145/1278280.1278289

Muender, T., Bonfert, M., Reinschluessel, A., Malaka, R., and Döring, T. (2022). "Haptic fidelity framework: defining the factors of realistic haptic feedback for virtual reality," in *Proceedings of the 2022 CHI conference on human factors in computing systems (CHI '22)* (New Orleans, LA, USA: Association for Computing Machinery). doi:10.1145/3491102.3501953

Pedersen, A. K., Isaksen, H., Riedel, M. B., Jørgensen, M., Paisa, R., and Nilsson, N. C. (2023). "Get a variable grip: A comparison of three gripping techniques for controller-based virtual reality," in 2nd IEEE VR 2023 Workshop on 3D Reconstruction, Digital Twinning, and Simulation for Virtual Experiences, Shanghai, China, 25-29 March 2023 (IEEE).

Quinn, P., Lee, S. C., Barnhart, M., and Zhai, S. (2019). "Active edge: designing squeeze gestures for the google pixel 2," in *Proceedings of the 2019 CHI conference on human factors in computing systems* (New York, NY, USA: Association for Computing Machinery). doi:10.1145/3290605.3300504

Ryu, N., Lee, W., Kim, M. J., and Bianchi, A. (2020). "ElaStick: A handheld variable stiffness display for rendering dynamic haptic response of flexible object," in *Proceedings of the 33rd annual ACM symposium on user interface software and technology* (New York, NY, USA: Association for Computing Machinery), 1035–1045.

Salazar, S. V., Pacchierotti, C., Tinguy, X. d., Maciel, A., and Marchal, M. (2020). Altering the stiffness, friction, and shape perception of tangible objects in virtual reality using wearable haptics. *IEEE Trans. Haptics* 13, 167–174. doi:10.1109/TOH.2020.2967389

Schmitz, M., Günther, S., Schön, D., and Müller, F. (2022). "Squeezy-feely: investigating lateral thumb-index pinching as an input modality," in *CHI conference on human factors in computing systems* (New York, NY, USA: Association for Computing Machinery). doi:10.1145/3491102.3501981

SenseGlove (2019). Getting started with the SenseGlove DK1. Avaliable At: https://www.senseglove.com/wp-content/uploads/2020/08/2020Manual1_3v1_non-booklet.pdf.

SenseGlove (2023). VR and haptic technology gloves. Avaliable At: https://www.senseglove.com/.

Tsai, H. R., Rekimoto, J., and Chen, B. Y. (2019). "ElasticVR: providing multilevel continuously-changing resistive force and instant impact using elasticity for VR," in *Proceedings of the 2019 CHI conference on human factors* *in computing systems* (New York, NY, USA: Association for Computing Machinery). doi:10.1145/3290605.3300450

Wang, D., Song, M., Naqash, A., Zheng, Y., Xu, W., and Zhang, Y. (2019). Toward whole-hand kinesthetic feedback: A survey of force feedback gloves. *IEEE Trans. Haptics* 12, 189–204. doi:10.1109/TOH.2018.2879812

Westling, G., and Johansson, R. S. (1984). Factors influencing the force control during precision grip. *Exp. Brain Res.* 53, 277–284. doi:10.1007/BF00238156

Whitmire, E., Benko, H., Holz, C., Ofek, E., and Sinclair, M. (2018). "Haptic revolver: touch, shear, texture, and shape rendering on a reconfigurable virtual reality controller," in *Proceedings of the 2018 CHI conference on human factors in computing systems* (New York, NY, USA: Association for Computing Machinery), 1–12. doi:10.1145/3173574.3173660

Witmer, B. G., and Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence teleoper. Virtual Environ.* 7, 225–240. doi:10.1162/105474698565686

Yabe, S. I., Kishino, H., Kimura, T., and Nojima, T. (2017). "Pseudo-haptic feedback on softness induced by squeezing action," in 2017 IEEE World Haptics Conference (WHC), Munich, Germany, 06-09 June 2017 (IEEE), 557–562. doi:10.1109/WHC.2017.7989962

Zhai, S., Milgram, P., and Buxton, W. (1996). "The influence of muscle groups on performance of multiple degree-of-freedom input," in *Proceedings of the SIGCHI conference on human factors in computing systems* (New York, NY, USA: Association for Computing Machinery), 308–315. doi:10.1145/238386.238534

Publication P5

Triggermuscle: Exploring Weight Perception for Virtual Reality Through Adaptive Trigger Resistance in a Haptic VR Controller

Carolin Stellmacher, Michael Bonfert, Ernst Kruijff, and Johannes Schöning

This publication investigates a sensory substitution approach with haptic stimuli from a controller's trigger button for weight perception. The VR controller *Triggermuscle* adjusts its trigger resistance according to the weight of a virtual object. Therefore, users need to adapt their index finger force to hold objects of different virtual weights. A spring mechanism enables dynamic and continuous adjustments. We conducted two user studies. The variations in trigger resistance were easily distinguished and associated with weight by some participants, while others did not notice them at all. We discuss how to overcome limitations and confounding factors in future research.

My contribution: I contributed to the discussion of the research idea (*conceptualization*), the study design (*methodology*), the literature review, and to writing and editing the manuscript (*writing – draft, review, & editing*). I supervised the project with JS (*supervision*).

Published in *Frontiers in Virtual Reality*, Section *Haptics*, Research Topic *Haptics and Human Interaction Model in Neurophysiology*, 2022 frontiers in Virtual Reality

ORIGINAL RESEARCH
published: 14 January 2022
doi: 10.3389/frvir.2021.754511

Check for updates

Triggermuscle: Exploring Weight Perception for Virtual Reality Through Adaptive Trigger Resistance in a Haptic VR Controller

Carolin Stellmacher[1], Michael Bonfert[1], Ernst Kruijff[2] and Johannes Schöning[3]*

[1]Faculty of Mathematics and Computer Science, University of Bremen, Bremen, Germany, [2]Institute of Visual Computing, Bonn-Rhein-Sieg University of Applied Sciences, Sankt Augustin, Germany, [3]School of Computer Science , University of St. Gallen, St. Gallen, Switzerland

It is challenging to provide users with a haptic weight sensation of virtual objects in VR since current consumer VR controllers and software-based approaches such as pseudo-haptics cannot render appropriate haptic stimuli. To overcome these limitations, we developed a haptic VR controller named *Triggermuscle* that adjusts its trigger resistance according to the weight of a virtual object. Therefore, users need to adapt their index finger force to grab objects of different virtual weights. Dynamic and continuous adjustment is enabled by a spring mechanism inside the casing of an HTC Vive controller. In two user studies, we explored the effect on weight perception and found large differences between participants for sensing change in trigger resistance and thus for discriminating virtual weights. The variations were easily distinguished and associated with weight by some participants while others did not notice them at all. We discuss possible limitations, confounding factors, how to overcome them in future research and the pros and cons of this novel technology.

Keywords: haptics, virtual reality, weight perception, adaptive trigger, controller design, psychophysics

Edited by:
Manivannan Muniyandi, Indian Institute of Technology Madras, India

Reviewed by:
Claudio Pacchierotti, Italian Institute of Technology (IIT), Italy
Gionata Salvietti, University of Siena, Italy
Ravali Gourishetti, Max Planck Institute for Intelligent Systems, Germany

****Correspondence:***
Carolin Stellmacher
cstellma@uni-bremen.de

Specialty section:
This article was submitted to Haptics, a section of the journal Frontiers in Virtual Reality

Received: *06 August 2021*
Accepted: *20 December 2021*
Published: *14 January 2022*

Citation:
Stellmacher C, Bonfert M, Kruijff E and Schöning J (2022) Triggermuscle: Exploring Weight Perception for Virtual Reality Through Adaptive Trigger Resistance in a Haptic VR Controller. Front. Virtual Real. 2:754511. doi: 10.3389/frvir.2021.754511

1 INTRODUCTION

Grabbing objects in reality provides humans with a haptic sensation of weight (Loomis and Lederman, 1986). Muscles, tendons and skin receptors sense the gravitational pull through proprioceptive and cutaneous stimuli (McCloskey, 1974; Brodie and Ross, 1984) and enable the natural perception of weights. Users of consumer virtual reality (VR) systems, however, cannot experience the same haptic weight sensation when grabbing virtual objects. Current consumer VR controllers are unable to render appropriate haptic feedback associated with weight, leaving users with identical haptic weight perception of virtual objects. In many cases, only visual cues are ambiguous and insufficient in conveying weight information. This limits a natural and realistic experience of weight in VR and results in a discrepancy between what users are familiar with from the real world and what they experience in virtual environments (VE).

Handheld controllers are typically used in consumer VR systems. With increasing consumer attention, these lightweight and mobile VR controllers become more relevant for designing interactions in VEs. They offer various components to register user input such as buttons, a trigger, or a trackpad, which provide users with different techniques to interact with the VE. However, to provide a haptic response during the interaction (e.g., the weight of a virtual object), current state-of-the-art VR controllers such as the HTC Vive or Oculus Touch controllers only offer vibrotactile rendering. This prevents users from having an appropriate haptic experience. It also rules out haptic feedback, which is specific to user input. To nevertheless enhance weight perception with consumer VR controllers and to overcome their current hardware limitations, pseudo-haptics were explored. This software-based approach manipulates the control/display (C/D) ratio between a user's hand movements and the rendered position in the VE during a lifting motion (Rietzler et al., 2018). This sensation of weight is caused by the user's arm movements being amplified when lifting heavier virtual objects with the controller. This effect was also observed when users lifted physical props in VR (Samad et al., 2019). While this method enables conveying a sense of weight in current consumer VR, it cannot resemble an actual haptic sensation of weight since users only receive visual feedback, but no haptic weight information.

Other researchers have addressed the hardware limitations of haptic weight rendering by proposing various lightweight and mobile VR devices. For such ungrounded haptic devices, it is particularly challenging to provide proprioceptive information as they cannot generate externally grounded forces that act on users. As an alternative, cutaneous stimuli are often used as a substitute to offset the lack of force feedback. For instance, a simulation of muscular grip forces is imitated through skin deformation at the finger pads of the index finger and thumb (Minamizawa et al., 2007a; Schorr and Okamura, 2017; Suchoski et al., 2018) or through skin stretch with asymmetric vibrations (Choi et al., 2017). Other wearable technology utilises electrical muscle stimulation to resemble a proprioceptive sensation by artificially pulling the user's arm downwards (Lopes et al., 2017). Other research has explored handheld devices with liquid-based haptic feedback to simulate the weight of fluid objects (Cheng et al., 2018), shape-changing abilities to provide haptic feedback through air resistance (Zenner and Krüger, 2019) or weight-shifting abilities (Zenner and Krüger, 2017) to generate haptic sensations for weight.

While these proposed haptic interfaces have succeeded in enhancing weight perception in VR, they are either designed for specific cases, rely on complex and expensive systems or need to be manufactured for different hand sizes. This makes them currently unsuitable for mass production and the consumer market. So far, no approach has considered established input components of current consumer VR controllers to render haptic stimuli for weight perception in VR.

In this paper, we propose an approach that—in contrast to related studies—utilises a standard button available in any consumer VR controller: the trigger. By varying the resistance of the trigger, which is normally constant by default, we extend the input component through output rendering to provide users with haptic feedback during the interaction. Hence, when users pull the trigger to grab a virtual object in VR, they need to scale their index finger force accordingly to the configured resistance displaying the virtual weight: The heavier a held virtual object is, the more finger force needs to be exerted onto the trigger.

We present *Triggermuscle*, a novel haptic controller that simulates the weight of virtual objects in VR through adjustable trigger resistance. As proof of concept, our system is built into the casing of an HTC Vive controller (see **Figure 1**). The novel spring mechanism is connected internally to the original trigger and dynamically modifies the trigger resistance according to the weight of the grabbed virtual object. A demonstration video of Triggermuscle is submitted as a **Supplementary Material**. The mechanism is built with inexpensive hardware components and can be easily tailored to other form factors due to the principle of force redirection, which demonstrates the potential for haptic weight rendering in different VR controllers. Additionally, as a handheld device, Triggermuscle fits a large range of users with various hand sizes. Enriching buttons with additional haptic feedback is increasingly evident in input device development. For instance, Sony recently released the DualSense controller for PlayStation 5 with actuated triggers (Sony, 2020), while Microsoft announced a locking feature for the triggers of the Xbox Elite controller (Microsoft, 2021).

What follows explores the capacity of variable trigger resistance to display virtual weight in VR. To do so, we

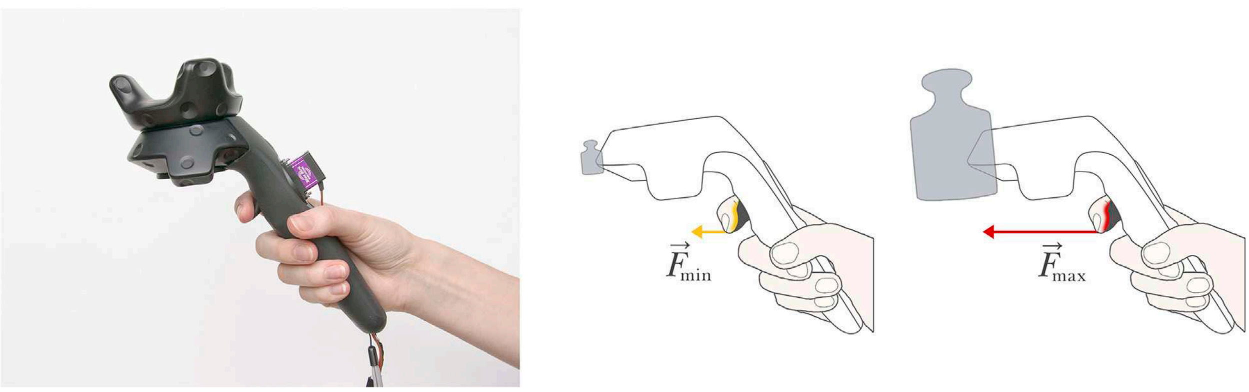


FIGURE 1 | (left) Our haptic controller Triggermuscle simulates the weight of virtual objects in VR through adapting the trigger resistance. The spring mechanism for a dynamic adjustment is built into the casing of an HTC Vive controller. An HTC Vive tracker is mounted to the top to enable spatial tracking. **(right)** Triggermuscle allows a continuous regulation of the trigger resistance according to the weight of the grabbed virtual object.

present the design and implementation of a preliminary prototype and of Triggermuscle as well as their different spring mechanisms, which we developed in an iterative and human-centred design process. In two user studies, we evaluate both controllers' hardware designs and investigate users' ability to discriminate different levels of resistance as well as the effect on weight perception in VR. Our research addresses two research questions:

RQ1: Do different trigger resistances influence the haptic perception of different virtual weights in VR?
RQ2: How can the intensity of the trigger resistance be quantified and mapped to convey distinguishable virtual weights?

We found large differences between participants for sensing change in trigger resistance and thus for discriminating different virtual weights based on the resistance. The variations were easily distinguished by some participants while others did not notice them at all. This points towards an association between trigger resistance and a sense of virtual weight in VR, but also towards confounding factors interfering with the perception of the trigger resistance. We discuss these issues to provide better insight into the problem space, to illustrate potential perceptual mechanisms that may have affected the experiment outcomes and to propose further studies and hardware designs based on our findings. We thus provide a novel hardware solution and, better understanding of the underlying (perceptual) mechanisms.

2 RELATED WORK

Extensive research in VR haptics has explored various strategies to enable experiencing physical properties of virtual objects during the interaction. This section presents an overview of grounded and handheld haptic technologies and emphasises approaches that convey a sense of weight in VR.

2.1 World- and Body-Grounded Haptic Devices

Humans feel an object's mass through the gravitational force pulling down the hand. To simulate this external pull in a force feedback display, the interface can be anchored in the environment or to the user's body. A widespread example for world-grounded interfaces is the Phantom Premium by 3D Systems (Massie and Salisbury, 1994; Systems, D., 2020), a 6 degrees-of-freedom (DoF) interface with a stylus as an effector mounted to the desk. A different approach anchors the force to the users' fingers via strings as done by the various SPIDAR interfaces (Sato, 2002) for simulating virtual weight in an early implementation (Ishii and Sato, 1994). Such wire-based force display systems were also adapted into body-grounded technologies to increase portability such as the HapticGEAR (Hirose et al., 2001) or Wireality for rendering complex shapes (Fang et al., 2020). Another wire-based variant anchors the user's hand to the respective upper arm (Tsai et al., 2019). By stretching the arm, the wire gets tightened and the user experiences multilevel resistive force and impact. Thanks to their mobility, these body-grounded interfaces allow users to move freely within the tracking area and are suitable if a larger space is required. At the same time, they typically need cumbersome and complex equipment that is time-consuming to set up.

Haptic interfaces worn on users' hands such as exoskeletons provide force feedback (Burdea et al., 1992; Bouzit et al., 2002; Ben-Tzvi and Ma, 2015; Gu et al., 2016; HaptX, 2020) that is grounded to different parts of the hand or arm (Nisar et al., 2019) and that can actively restrict the finger movement with motors and complex mechanics. This can make such devices bulky, tethered, expensive or limit hand flexibility. In contrast, gloves (Giannopoulos et al., 2012; Martínez et al., 2016; Marquardt et al., 2018; Manus, 2020) typically do not use motors and can, therefore, overcome shortcomings related to the actuation. Gloves can also track users' fingers and provide vibrotactile feedback or other cutaneous stimuli. Researchers rendered various physical properties of virtual objects including shape (Solazzi et al., 2007), contact forces (Leonardis et al., 2015), texture (Gabardi et al., 2016), or inertia (Girard et al., 2016). Nonetheless, putting on gloves or hand-mounted equipment like exoskeletons and thimbles can be time-consuming and requires hygienic considerations. Further, they need to be adjustable or manufactured in different sizes to fit a diverse range of users.

2.2 Handheld Haptic Devices

In contrast to such wearable haptic interfaces, handheld haptic devices are ready for use when picked up. They do not physically restrict the user's movements and flexibly fit a large range of hand sizes. Most current consumer VR systems include such handheld controllers by default. In recent years, the development of handheld haptic devices that generate physical forces for haptic feedback as well as controller-based interaction techniques has received considerable attention. For instance, the CapstanCrunch allows to feel rigid and compliant objects (Sinclair et al., 2019), the TORC creates a haptic sensation for texture and compliance (Lee et al., 2019) and the controllers NormalTouch and TextureTouch render shape and texture through tilting a platform at the user's index finger pad (Benko et al., 2016). PaCaPa is a haptic display for an object's size, shape and stiffness by tilting movable wings, but cannot render resistance (Sun et al., 2019). Haptic Links generates resistance by mechanically constraining the relative movement between two controllers (Strasnick et al., 2018), while Thor's Hammer creates force feedback through airflow with propellers (Heo et al., 2018). The CLAW is a handheld device that integrates multiple haptic technologies to simulate a range of haptic sensations (Choi et al., 2018). It renders kinaesthetic forces at the index finger during grasping and touching which allows feeling the shape and stiffness of virtual objects. Additionally, a voice coil actuator produces vibrations for different surface textures. Another interface for displaying surface properties is the Haptic Revolver (Whitmire et al., 2018). Shapes and shear forces that occur when gliding along a surface are rendered at the fingertip by rotating a wheel with a direct current (DC) motor. The haptic wheels are customisable and can provide various textures and shapes, but can also comprise active electronic components such as buttons, switches and joysticks. Transcalibur enables shape perception through inertia (Shigeyama et al., 2019). Variable weight distributions along the controller are realised with shifting weights. Moving the controller through space makes the inertia noticeable for users and creates a haptic shape illusion. A similar concept was proposed with ShapeSense. Movable surface elements increase or decrease the surface area of the controller (Liu et al., 2019). While these mechanisms can render the distribution of an object's mass, they cannot render the absolute mass of objects. Furthermore, even though the proposed devices allow haptic rendering for various object properties, they do not address the weight of virtual objects.

2.3 Haptic Devices for Weight Simulation

To provide users with a weight sensation during the interaction with virtual objects in VR, previous work explored diverse approaches. For example, electrical muscle stimulation was used to induce contractions of the user's muscles while lifting virtual objects. The system actuates the user's triceps to simulate the weight of a held virtual object by inducing a downward movement of the arms (Lopes et al., 2017). Another concept of inducing a sense of weight for VR applications is skin deformation on the finger pad. This imitates the stretch of the skin from the downward pull of the object's surface. One way to do so is using small actuated belts that are strapped around users' finger pads which has been shown to generate a reliable weight sensation (Minamizawa et al., 2007b). The refined implementation of this approach induced the impression of grip force, gravity and inertia by stretching the skin on the finger pad with the attached belt, without the need for proprioceptive sensations (Minamizawa et al., 2007a). This type of approach was also explored in the context of augmented reality (AR) rendering weight and shear forces (Scheggi et al., 2010). To account for the combination of physical and virtual objects in AR, another implementation placed the finger-worn device as a ring around users' fingers leaving the finger pads free for the interaction with physical objects (Pacchierotti et al., 2016; Maisto et al., 2017). Other haptic devices use actuated plates to achieve skin stretch such as a finger-worn device that slides the contact area at the user's index finger pad to mimic weight and friction (Kurita et al., 2011). Further, scaling inertial forces rendered with a 3DoF wearable device on the finger pad showed an increase of the perceived weight of an object moved by the user (Suchoski et al., 2018). A different implementation demonstrates a handheld controller with movable plates in its handle to resemble the friction between an object and the hand during grasping (Provancher, 2014). Another finger-mounted device is Grabity, which simulates grip forces and a sensation of weight. The device is mounted on the thumb, index and middle finger and applies kinaesthetic forces for rendering shape. To render weight, asymmetric vibrations of voice coil actuators stretch the skin at the finger pad resembling the pull of gravity. The participants in the evaluation successfully distinguished the objects of different weight but felt the vibration cues even stronger than the weight cues (Choi et al., 2017).

The handheld VR controller Drag:on adjusts its surface area to generate varying haptic sensations for experiencing drag and weight. As the concept depends on air resistance, the different object properties are only noticeable when the device is moved through space. Due to the flat controller design built with fans, the effect is dependent on the orientation of the controller (Zenner and Krüger, 2019). The same authors also created Shifty, which enhances the perception of the dimensions of virtual objects by changing the controller's weight distribution. An internal weight is moved along the longitudinal axis shifting the centre of mass away from the hand. This increases the leverage and therefore feels like holding a heavier object (Zenner and Krüger, 2017). An increase in the possible rendered shift was achieved through combining the haptic device with haptic retargeting (Zenner et al., 2021). Rendering shifting weights on a 2D plane was also achieved in a handheld controller using jet propellers. Aero-plane generates force feedback with up to 14 N that can be interpreted by the user as weight while holding the device level. This was found to increase the perceived immersion and realism (Je et al., 2019). Finally, in GravityCup, the actual weight of the device changes. It is filled with water or emptied again to render inertia and weight of liquids. The user holds the interface by a handle like a cup. The interface requires a separate wearable bag with water to fill the haptic display as needed (Cheng et al., 2018). So far, these proposed devices often rely on complex hardware, might feel cumbersome to users or target specific use cases which limits their use for haptic weight rendering in commercial VR.

2.4 Software-Based Approaches for Weight Simulation

As discussed, haptic devices are limited in their application by a number of factors. Moreover, one additional constraint is the availability of the hardware. Beyond technical feasibility, a haptic display needs to be universal, flexible and affordable enough to be established as a standard interface in VR interaction. To overcome hardware limitations and deliver haptic experiences readily available to users, researchers proposed various software solutions for pseudo-haptics. This term describes haptic illusions through visual, auditory or multimodal stimuli without actual touch.

In terms of weight perception, the manipulation of the C/D ratio between users' hand movements and the rendered position in VR has proven effective. It was demonstrated in a non-VR setup that this mismatch strongly influences the perception of mass (Dominjon et al., 2005). The effect has been replicated successfully in VR. In an experiment, participants lifted two physical boxes with their hands. An increase in the offset for heavier virtual boxes resulted in an amplification of users' hand movements and a heavier perceived weight (Samad et al., 2019). This method was also applied to the interaction with a consumer VR controller and produced corresponding results (Rietzler et al., 2018).

Analogously to the modification of the translational C/D ratio, another approach changed the rotational C/D ratio depending on an object's weight, thus, the rotational motion is scaled relative to the mass of the object. A user study confirmed that this method effectively, realistically and robustly conveys different weights. At the same time, it does not compromise the perceived controllability. Furthermore, the authors proposed the manipulation of the pivot point during rotation and the scaling of rotational motion to convey the distribution of mass within an object (Yu and Bowman, 2020).

Such software-based approaches have been shown to provide users with an experience of virtual weight, but they cannot render actual force stimuli. This limits their ability to haptically convey weight in VR. Unlike previous devices or software-based approaches, Triggermuscle offers a novel hardware solution built into a commercial handheld VR controller to enhance weight perception during the interaction with a virtual object in VR. With our technology, we extend the trigger's capabilities towards generating haptic feedback and explore the effect of adaptive trigger resistance on the perception of virtual weight in VR.

3 ADAPTIVE TRIGGER

This section describes the background in haptic weight perception during grasping and how we addressed this in the concept and the first implementation of the adaptive trigger inside a prototype. The section concludes with the evaluation of the prototype in a pilot study and the findings that influenced the development of Triggermuscle.

3.1 Background in Weight Perception

Our adaptive trigger is informed by humans' perception of weight cues through the haptic sense (Loomis and Lederman, 1986). In addition, to grasp and lift an object the human brain initially incorporates visual cues (Gordon et al., 1991) and previous lifting experiences (Van Polanen and Davare, 2015) to predict an object's weight and scale finger forces accordingly. Touching and lifting the object then supplies simultaneous haptic cues obtained from cutaneous stimuli registered by receptors in the skin and proprioceptive stimuli obtained from muscles and tendons (McCloskey, 1974; Brodie and Ross, 1984). Depending on the updated weight perception, enough grip force is applied to overcome the gravitational pull, but at the same time causing no damage to the object (Westling and Johansson, 1984). The result is a direct relationship between the physical weight and the applied grip forces: The heavier the object, the more manual force needs to be applied. This principle forms the main instigator for our hardware design. Increasing the grip force to a sufficient amount is enabled by isometric contractions in the muscles of the hand and arm (Johansson and Westling, 1988), meaning the muscle tension is adjusted accordingly to the weight, but no muscle movement takes place. When lifting the object away from the supporting surface, the contractions switch to isotonic which keeps the muscle tension static while the length of the muscle changes, e.g., to flex the elbow.

Grip forces are not only scaled according to the gravitational pull, but are also influenced by various tactile cues derived from material properties such as surface texture (surface-weight illusion) (Johansson and Westling, 1984; Flanagan et al., 1995) or material (material-weight illusion) (Ellis and Lederman, 1999) and spatial properties such as shape (shape-weight illusion) (Jenmalm and Johansson, 1997) or size (size-weight illusion) (Ellis and Lederman, 1993). These studies have shown that the illusions provoke modulated grip forces and influence the perceived weight. For instance, an increased grip force due to smoother surface texture leads to higher perceived weight (Johansson and Westling, 1984; Flanagan et al., 1995). In such cases, cutaneous receptors detect less frictional force between the skin and the object's surface leading to higher grip forces to prevent the object from slipping. Based on their findings, the researchers argued that "grip force may be a useful cue for discriminating weight" (Flanagan et al., 1995).

Such haptic illusions caused by stimuli unrelated to gravity could contribute to a successful substitution of haptic weight cues occurring during the grip. The sensory substitution implies that haptic stimuli are registered through another sense as they normally are or at a different location (Kaczmarek et al., 1991). This is necessary for most haptic interfaces to compensate for the lack of corresponding physical stimulation when virtually interacting with an object. Especially handheld VR devices displaying virtual weight rely on substitutional stimuli to compensate for the lack of gravitational force pulling down the user's hand. Simulating this force has been previously done, e.g., by deforming the user's skin at the fingertips through stretching the skin or through asymmetric vibrations (Choi et al., 2017). Moreover, as haptic stimuli unrelated to gravity such as surface texture or material have been shown to create weight illusions and

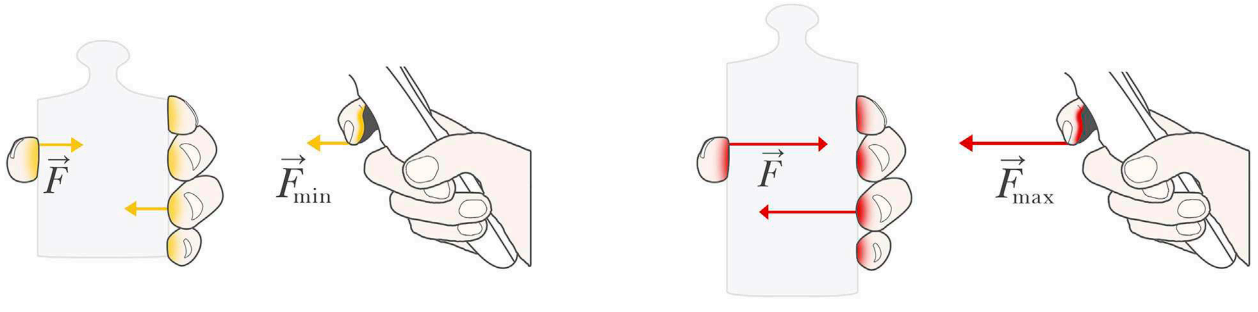


FIGURE 2 | Modulated finger forces from reality occurring during the grasping of objects with different weights are translated into the resistance of the trigger.

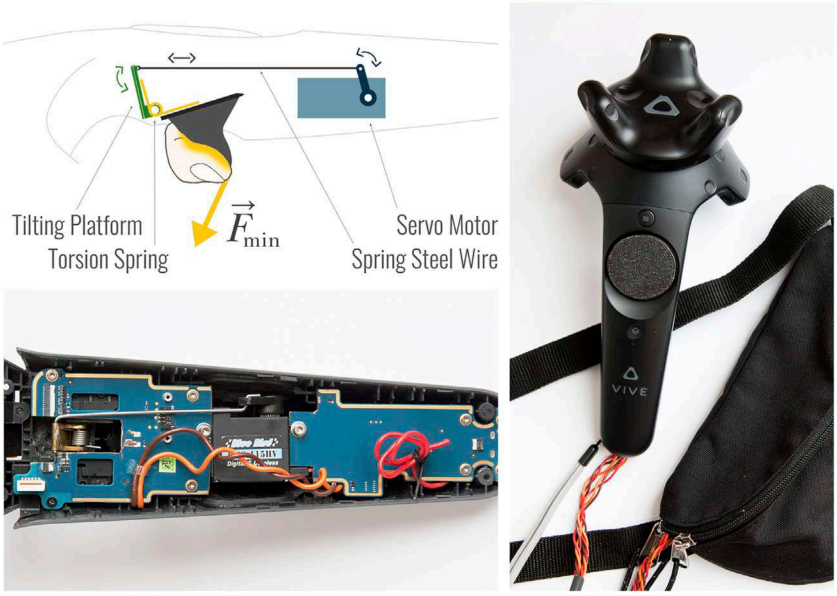


FIGURE 3 | (top left) Schematic illustration of the spring mechanism of our prototype utilising a torsion spring. **(bottom left)** Spring mechanism embedded into the casing of an HTC Vive controller. **(right)** Our prototype for adaptive trigger resistance. Electrical components are connected via cable to the prototype's bottom and carried in a small bag. An HTC Vive tracker attached on top ensures the spatial tracking of the device.

induce a modulation of grip forces, we assume that providing variable resistances as a haptic stimulus at the users' index fingers and thereby provoking a modulation of grip forces might enable a haptic weight perception in VR.

3.2 Concept

The concept of our adaptive trigger incorporates the previously described relation between grip forces and the perceived weight and transfers it to the established interaction technique of any consumer VR controller. With such devices, grabbing a virtual object typically involves pulling the trigger, which requires muscle force of users' index fingers to overcome the constant resistance. By adjusting this resisting force, substitutional stimuli (Kaczmarek et al., 1991) are displayed as weight cues and users need to adjust their index finger forces according to the weight of the grabbed virtual object. For example, the heavier the virtual object, the stronger the trigger must be pulled. An illustration of the intended effect is shown in **Figure 2**. Early evidence has demonstrated that the pull of the trigger can be interpreted as varying the grip force that the user exerts onto a virtual object enabling loose and firm grasping for controlling the object's rotation (Bonfert et al., 2019). As the grip force required for holding an object correlates with its weight, higher trigger resistance provoking higher index finger forces might consequentially be interpreted as increased weight.

While our adaptive trigger transfers the haptic recognition of weight onto a one-finger interaction, other haptic interfaces have demonstrated that rendering haptic stimuli only at users' index fingers can be sufficient to enhance haptic object perception in VR (Benko et al., 2016; Choi et al., 2018; Whitmire et al., 2018). In particular, applying resistive forces to restrict the index finger's movement during grasping has been shown to enhance the perception of rigid and compliant objects (Sinclair et al., 2019). With Sony and Microsoft incorporating haptic feedback into their triggers for game experience, the future availability thereof is another strong argument for exploring the potential of triggers with adaptive resistance for the perception of various haptic events, in our case weight perception.

We implemented our adaptive trigger in an iterative process following a human-centred design approach. First, we built a prototype of a spring mechanism to dynamically adjust the trigger resistance which we present in **section 3.3**. We evaluated the effectiveness of the resistance range and the technical implementation in a pilot study. Based on our findings, we revised the spring mechanism and built the improved haptic controller Triggermuscle, shown in **section 4**.

3.3 Prototype

The first implementation of our adaptive trigger is based on the typical construction of standard triggers: Pulling the trigger compresses one leg of a torsion spring whereby its tension exerts a force in the opposite direction, i.e., resisting the finger's pull. To establish a change in tension force, we constructed a mechanism that rotates the spring's second leg before a pull motion, increasing or decreasing its angle. The resulting adjustment of the trigger resistance is dynamically performed by a high-voltage (7.4 V) digital micro servo (BMS-115HV) which rotates the usually fixed leg via a connected tilting platform. The entire mechanism is built into the casing of an HTC Vive controller and connected to the original trigger, as illustrated in **Figure 3**. In contrast to the later Triggermuscle controller, the prototype fully accommodates the servo inside the casing.

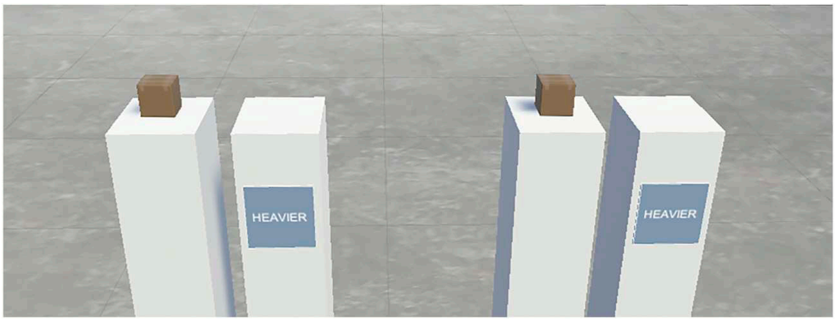


FIGURE 4 | Setup in the VE used for the main user study. Both boxes had to be lifted and placed onto the platform right next to it. Virtual "HEAVIER"-buttons on both target platforms allowed participants to log in their response. The same setup was used for the preliminary testing of the pilot study.

With our used spring model, the prototype renders a continuous range of resistance between 19.27 Nmm and 47.61 Nmm with a fully pulled trigger. Beyond that, users could increase the finger force further, but without moving the trigger or changing the input, identical to a conventional trigger. The chosen spring model offered the greatest possible resistance range when installed inside the mechanism and its level of resistance was closely located to the one of an original Vive controller. Informal testing in the lab with three users suggested comparability between the middle value of the prototype's range and the trigger resistance of the original HTC Vive controller. The resistance values are calculated based on the path-force ratio of the spring model (29.44 Nmm/103.89°), the respective compression angle computationally set by the servo, the additional 18° compression when the trigger is pulled and the spring's preloaded angle of 50° when installed inside the mechanism. The latter two angles were carefully measured manually using visual scales. The prototype's total resistance range achieves a maximum increase of 147%. Humans are known to perceive a difference in spring stiffness between 15 and 22%, also known as the Weber Fraction (WF) (Jones and Tan, 2013).

Apart from the modified resistance, the haptic sensation of pulling the trigger is maintained, including the final *click*. This occurs when users fully pull the trigger which then mechanically pushes the original mini button that registers the digital signal at the maximum limit. The signal is send via cable to an ESP32 microcontroller unit (MCU) which also drives the servo and communicates with Unity 2018.3 via Bluetooth. Along with a 11.1 V lithium polymer battery and a battery eliminator circuit (BEC) component, the MCU is carried in a small bag on the user's back and connected to the controller's bottom via cable. Since the original tracking components were removed, an HTC Vive tracker 1.0 is mounted to the top of the controller.

3.4 Pilot Study

Our pilot study evaluated users' ability to perceive and discriminate different resistances in VR while using the prototype. Similar to previous haptic device research (Dominjon et al., 2005; Maereg et al., 2017; Suchoski et al., 2018; Ryu et al., 2020), we conducted a psychophysical experiment to measure the just noticeable difference (JND) of adaptive trigger resistance in VR. In addition, we also carried out semi-structured interviews to qualitatively assess subjective perception. In accordance with our research goal, to explore weight perception through adaptive trigger resistance, our initial objective also included the prototype's effect on the perception of virtual weight. However, preliminary testing, in which eight participants lifted and compared the weight of two boxes in VR (see **Figure 4** for the VE), suggested that weight perception was not influenced by the varying intensity of the trigger resistance. Based on participants' reports, we assumed that the visual modality of the box-lifting task dominated the perception. To address this issue, we simplified the visual input and focused, as a first step in the pilot study, only on users' ability to discriminate different trigger resistances, without additional weight perception. The simplified pilot study is described below.

3.4.1 Experiment

We recruited nine participants (two females, seven males) aged 21–50 (M = 28, SD = 9.22), of which the majority (7) reported previous VR experience. One participant did not produce valid data and was excluded from further analysis. At the beginning of the experiment, participants gave their consent to take part in the study and were not made aware of the altered trigger resistance to ensure unbiased experience. Throughout the experiment, participants used the prototype and wore the head-mounted display (HMD) of the HTC Vive system. Noise-cancelling headphones played neutral music to block the motion noise of the servo and to avoid possible bias.

To implement an interaction task that involved pulling the trigger, participants were asked to change the colour of a grey virtual wall in VR by pressing the button. In each trial, participants consecutively activated two colours, magenta or green. For each activation, a different trigger resistance was rendered. Each colour was then deactivated, returning the wall to grey, as soon as the trigger was released. At the end of each trial, participants chose that colour that felt heavier to activate and logged their response by touching the virtual interface button in the respective colour. To assess participants' ability to discriminate different resistances, we followed the method of constant stimuli with a two-alternative forced choice (2AFC) paradigm (Jones and Tan, 2013), as this is said to produce more accurate results than alternative methods (Simpson, 1988; Guilford, 1955). Each activation was haptically rendered with different trigger resistances using the prototype described in **section 3.3**. In each trial, participants were presented with the same standard resistance (19.27 Nmm) and with one of four preselected comparison resistances (26.35, 33.44, 40.52, 47.61 Nmm). These were equally distributed along the prototype's resistance range. The comparison resistances were computed based on the compression angles set by the servo and by the path-force ratio of our used torsion spring model. Typically, the comparison values are spaced on either side of the standard value (Jones and Tan, 2013). However, we were concerned that presenting only half of the range as the maximum resistance change might be too subtle to be noticed by participants. We, therefore, chose the minimum resistance as the standard value, similar to (Maiero et al., 2019). Each comparison stimulus was tested 10 times, resulting in a total of 40 trials. The order of all trials and the appearance of the standard and comparison resistance as well as the appearance of green and magenta within one trial were randomised. Three pre-task trials allowed participants to familiarise themselves with the procedure. Upon task completion, we carried out semi-structured interviews to assess participants' self-reported experience.

3.4.2 Results

We measured the proportions of "heavier"-responses for each tested comparison resistance, plotted psychometric functions (PFs) and assessed the goodness-of-fit (Schütt et al., 2016) using the MATLAB toolbox psignifit 4 (Schütt, 2019). Based on the results, only three participants qualified for further JND computation, the point of subjective equality (PSE) and the WF. Three of the excluded data sets performed around the guess rate. The other two performed almost perfectly, not allowing an assessment of discrimination sensitivity, since the method of constant stimuli of psychophysical testing requires a decreasing range of correct responses between 100 and 50% in the case of our 2AFC task. For the remaining three participants, the average JND was 3.90 Nmm ($SD = 0.79$), resulting in an average WF of 14.70% ($SD = 3.60$). This level of sensory precision is slightly below the previously mentioned 15–22% WF in the literature of spring stiffness discrimination. However, due to the small number of considered data sets, the results should be treated with caution.

3.4.3 Discussion and Implications

At this point, the results remain inconclusive if the adaptive trigger resistance can be discriminated sufficiently in VR. Nonetheless, our findings offer the incentive to continue exploring the adaptive trigger resistance since the data sets of three participants suggest that their identification of the heavier activation was influenced by the intensity of the resistance. Further, two participants achieved an almost always perfect identification which further indicates an influence of the adaptive trigger resistance on their perception. Our findings also highlight two key limitations with the prototype's haptic feedback which we will discuss in the following.

Simplifying the visual input after the preliminary testing showed an improvement in participants' perception of the adaptive trigger resistance. We assume it enabled a shift in their attention and allowed participants to detect the provided haptic feedback by themselves, as they were kept unaware of it. This assumption is in line with other haptic VR research also observing a domination of the haptic sense by the visual sense (Ban et al., 2012; Azmandian et al., 2016; Choi et al., 2018; Degraen et al., 2019). Providing stronger haptic stimuli in future investigation might, therefore, achieve a more balanced perception of both senses. Consecutively, this might allow more users to sense the resistance as well as sense the resistance more strongly. To continue the exploration and investigate whether higher resisting forces improve the sensing of the adaptive trigger resistance and if they can influence the perception of weight in VR, we plan to establish a wider resistance range.

Furthermore, statements obtained from the interview showed that a subtle vibration occurring as a side effect during the servo's adjustment was noticed by all participants. This suggests another possible diversion of participants' haptic attention, additionally preventing them from recognising the resistance change. This could have again been amplified by the fact that participants did not know about the adaptive trigger resistance. We plan to address the vibration as a possible side effect through modifications in the hardware as well as software design. Modifying the hardware design by increasing the distance between the servo and participants' hands, adding additional damping or by using a different type of actuation might help to reduce the exposure to the servo's subtle adjustment vibrations. Implications for the software design are derived from our observations during the preliminary testing with the box-lifting task. The implemented software for that task adjusted the servo angle when participants reached for the virtual box and intersected its collider with the one from the virtual controller. To decouple the controller's adjustment from the moment of participants preparing to pull the trigger, we plan to set the servo's angle independent from the lifting motion.

4 TRIGGERMUSCLE

To overcome the drawbacks of the prototype, we built Triggermuscle, which is shown in **Figure 1**. The key change implemented in Triggermuscle is the revised spring mechanism that utilises an extension spring. This allows a larger manipulation of the exerted force in contrast to the torsion spring used for the prototype. Therefore, pulling the trigger of Triggermuscle stretches the attached extension spring and makes the exerted force noticeable to users' index fingers as the trigger resistance. Thus, changing the length of the extension spring enables the adjustment of the trigger resistance.

4.1 Implementation

The revised spring mechanism is again built into the casing of an HTC Vive controller and attached to the original trigger, as illustrated in **Figure 5**. The dynamic adjustment of the extension spring is established with a high-voltage (6.0 V)

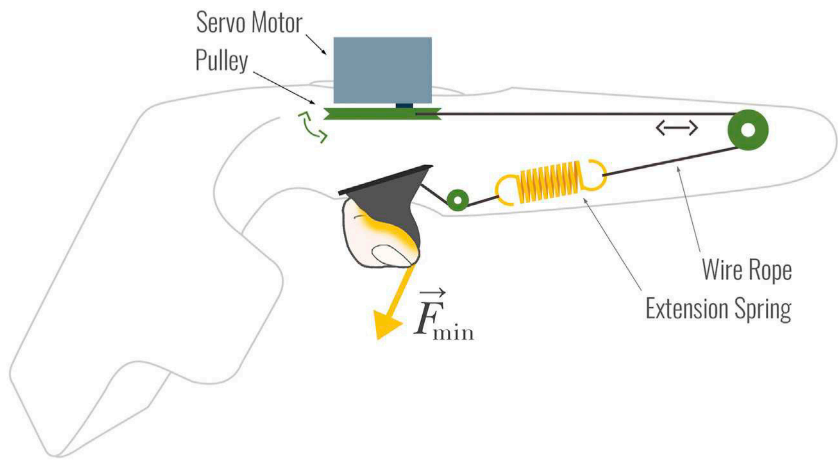


FIGURE 5 | Schematic illustration of Triggermuscle's spring mechanism utilising an extension spring.

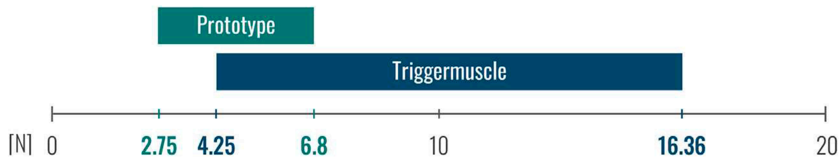


FIGURE 6 | Both ranges of trigger resistance of the prototype and Triggermuscle. The torsion spring's torque values [Nmm] of the prototype were converted into [N] for the purpose of comparability.

digital micro servo (BMS-210DMH). However, in contrast to the prototype, the servo was moved upwards and installed into the original trackpad component of the controller. The revised servo position increases the distance to the user's hand and was intended to reduce the sensed vibrations reported in the pilot study. To provide further damping, a protective silicon cover for the HTC Vive controller was wrapped around the handle. With an attached pulley on top, the servo is connected to the extension spring via a thin 1 mm wire rope. Despite its small diameter, the wire rope is strong enough to handle the forces and inelastic to ensure accurate translation. Changing the servo's angle rotates the attached pulley and winds the connected wire rope, thus pulling or releasing the spring. A second, smaller redirect pulley in the controller's bottom secures a stable guiding of the wire rope. Due to this principle of force redirect, the mechanism can be tailored to other form factors for different interaction devices.

The dimensions of the controller casing allow a spring stretch of up to 20.3 mm. For the selection of the suitable spring model, the same selection criteria were applied for Triggermuscle as for the prototype. In combination with the used extension spring model (spring rate of R = 0.592 N/mm, minimum force 1.33 N), a continuous regulation of the resistance is achieved from approximately 4.29–16.36 N with a fully pulled trigger (5 mm stretch adding 2.96 N). Increasing the finger force further beyond the set maximum limit does not move the trigger or change the input, identical to a conventional trigger. Note that we use the force unit [N] for the extension spring which differs from the unit [Nmm] for the torsion spring torque of the prototype to emphasise the correct units of the spring types and for the readers to easily identify the prototype and Triggermuscle. The total resistance range illustrates an increase of over 281%. In contrast, the prototype allowed a smaller increase of 147%. Triggermuscle, therefore, exceeds the previously tested range of trigger resistance of the prototype by 134%. For the purpose of comparison, both ranges are illustrated in **Figure 6** in which the resistance values of the prototype were converted into [N].

Similar to the prototype, the haptic sensation of pulling the trigger is maintained, apart from the modified resistance. The MCU, the lithium polymer battery and the BEC are again carried in the bag previously used for the prototype. The final controller is shown in **Figure 7**. An HTC Vive tracker 1.0 attached to the controller's top provides spatial tracking since the original tracking components were removed from the casing. The total weight of Triggermuscle including the HTC Vive tracker 1.0 (90 g) and its mounting (10 g) is 300 g. In comparison, the original HTC Vive controller weighs 200 g. The bag carrying the battery (250 g), MCU (10 g) and BEC (22 g) weighs 350 g. Further details are described in the workshop paper (Stellmacher, 2021).

5 STUDY

Our main user study evaluated the revised technical implementation of the adaptive trigger that is built into

FIGURE 7 | (left) Our haptic VR controller Triggermuscle. **(top right)** Attachment of the HTC Vive tracker to enable spatial tracking. **(right bottom)** Electronic components (MCU, the battery and the BEC) are carried in a small bag.

Triggermuscle. Our main objective was to explore whether the increased resistance range makes more users notice the change in intensity. In addition, we investigated whether differently intense trigger resistances resemble a perception of virtual weight in VR. To do so, this study repeated the psychophysical experiment of our pilot study by using the method of constant stimuli. However, we adapted the interaction task to the main goal of the user study and repeated the box-lifting task used in the preliminary testing of the pilot study.

5.1 Participants

We recruited 21 participants (five females, 16 males) aged 19–29 (M = 22.67, SD = 2.78). Most of them (19) stated previous VR experience, more than half (16) were familiar with non-VR game controllers. No participant was previously involved in the pilot study.

5.2 Task and Stimuli

Participants lifted and compared two visually identical boxes and identified the heavier one. The setup in the VE is shown in **Figure 4**. The virtual weight of both boxes was haptically rendered with different trigger resistances. In each trial, participants were presented with the same standard resistance and with one of five preselected comparison resistances. The value of the standard resistance was 4.29 N (0% of the range). The five comparison resistances were 4.46 N (2%), 4.79 N (5%), 6.09 N (19%), 8.67 N (46%) and 13.82 N (100%), with each being repeated ten times. Therefore, a

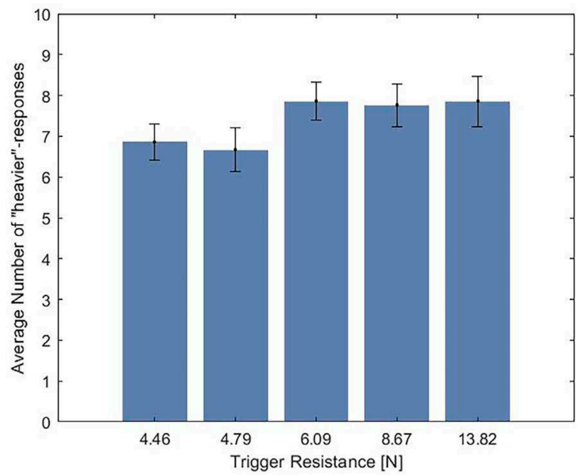


FIGURE 8 | Average number of "heavier"-responses for each tested resistance intensity from our main user study with error bars representing the standard errors. The maximum possible value is ten due to the total amount of ten trials per intensity level. A decrease was expected for the number of "heavier"-responses over the decreasing range of the resistance intensity, which, however, cannot be observed in the average responses.

total of 50 trials were conducted. The order of all trials and the appearance of the standard and comparison resistance within one trial was randomised. The possible maximum value of 16.36 N of Triggermuscle was restricted to 13.82 N for this user study to avoid wearing out the

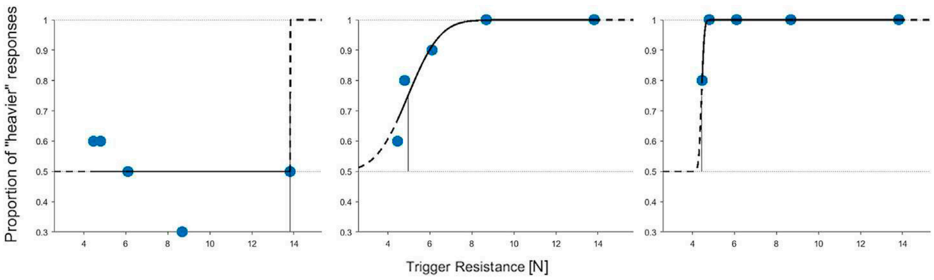


FIGURE 9 | Psychometric functions of three participants representing three commonly observed behaviours in our user study for the perception of Triggermuscle's adaptive trigger resistance. **(left)** "Heavier"-responses are located closely around and below guess rate not allowing a direct conclusion for the perception of different resistances. **(middle)** Proportion of "heavier"-responses decreases with the decrease of resistance intensity. This perception shows the full range of discrimination sensitivity of the sensory system. **(right)** High proportion of "heavier"-responses even with smaller resistances.

components. The lower half of the tested resistance range was covered by four comparison values since we considered the highest resistance to be easily recognised. We also assumed that participants would easily become aware of the change in the haptic feedback and consequently be more attentive to smaller changes. We also expected our sample to include participants who almost always successfully identify the heavier resistance as observed in two participants in the pilot study, and thus motivating the small 2% value. The crooked percentage values are due to the resolution of the servo angles.

5.3 Procedure

At the beginning of the task, participants received instructions but were kept unaware of the adaptive trigger resistance and possible vibrations as a side effect. They performed three pre-task trials in the VE to familiarise themselves with the procedure of lifting the boxes and selecting the heavier one. Three comparison resistances were rendered during those trials: 7.29 N (25% of the range), 7.72 N (29%) and 8.15 N (32%). Since we considered higher resistances of Triggermuscle to be easily recognised, the values were distributed around the lower third of the range to ensure an unbiased starting position for the psychophysical testing during the experiment task. To avoid possible influences of the servo's motion noise, participants wore noise-cancelling headphones and listened to neutral music. After completing the experiment task, we carried out semi-structured interviews to assess the self-reported experience of participants.

6 RESULTS

We assessed the influence of adaptive trigger resistance on the perception of virtual weight based on quantitative data recorded during the task and qualitative data from the interview.

6.1 Trigger Resistance Discrimination

The average number of "heavier"-responses for each tested resistance intensity are shown in **Figure 8**. We expected the number of "heavier"-responses to clearly decrease with decreasing resistance intensity. This behaviour, however, could not be observed in the average responses. For a more differentiated assessment, PFs for all 21 participants were fitted, again using the MATLAB toolbox psignifit 4 (Schütt, 2019). We also reassessed goodness-of-fit based on the calculated deviance to assess the proximity between the fitted dataset and the underlying model. It asymptotically converges to 1.0. However, the lower the mean proportion of correct responses, the higher the expected deviance is. A "typical cut off [is] around the value 2 for what is often regarded as a still" well behaved" data set" (Schütt et al., 2016).

TABLE 1 | Results of four participants of our psychophysical user study.

Participant	JND [N]	WF [%]	PSE [N]
5	0.81	16.27	4.96
13	0.06	1.25	4.46
17	0.13	2.23	5.72
18	0.13	2.14	6.11
Mean	0.28	5.47	5.31
SD	0.35	7.21	0.74

The results demonstrate highly diverse subjective perception of trigger resistance. The observed perception can be categorised into three behavioural patterns based on their deviance values and average percentages of "heavier"-responses. Each category is represented by one of the three exemplary PFs (see **Figure 9**). Seven PFs are located at the lower end of the spectrum (represented by **Figure 9** (left)). The average percentage was of 50.3% (SD = 5.50) for the "heavier"-responses with a deviance often well above 2.00. Since these responses are located closely around the guess rate of 50% (in our 2AFC case), not enough evidence exists to support our assumption that they were influenced by the intensity of the trigger resistance. The second category includes four data sets that exhibit the expected decrease in discrimination ability with a decrease in resistance level (represented by **Figure 9** (middle)). Another seven PFs are located at the higher end (represented by **Figure 9** (right)). They show an average success rate of

94.30% (SD = 1.80). Five exhibited a deviance well above 2.00, two a value of 0.00 and 0.02, respectively. These participants almost always perfectly identified the heavier box, even with smaller resistances, and did not produce data within a resistance range where their perception changes according to the level of resistance. Three PFs showed a deviance above 2.0, indicating a high discrepancy from the underlying model. Since the method of constant stimuli of psychophysical testing requires a decreasing range of correct responses between 100 and 50% in the case of our 2AFC task, only the four data sets represented by **Figure 9** (middle) qualify for JND calculation. The remaining 17 data sets did not produce data that would allow understanding the discrimination sensitivity of the sensory system around the threshold and the resulting perception. Most of these participants are associated with both ends of the spectrum depicted in **Figure 9**. The results of the four qualified participants for JND, WF and PSE are listed in **Table 1**. Their sensory precision was determined with an average JND of 0.28 N (SD = 0.35). This resulted in an average WF of 5.47% (SD = 7.21), which is below the reported WF of 15–22% in the literature of spring stiffness discrimination. However, due to the small number of included data sets these values can only be interpreted as a first indicator of future studies. Nonetheless, the low average is caused by three out of four participants exhibiting a WF of equal or below 2.23%. Only one participant produced a WF of 16.27%, which is in line with the literature.

6.2 Interview Feedback

Statements obtained in the interview revealed that one-third of participants self-reported the adaptive trigger resistance as the cue for identifying the heavier box. Two of them additionally incorporated subtle vibrations which occur as a side effect during the servo's adjustment. The perception of virtual weight was described through "the trigger was harder to press for heavier boxes" or "depending on how much I had to press the index finger". Lifting the boxes was described as "heavier or less heavy", lifting heavier boxes was characterised as "more demanding" and one participant reported that grabbing a heavier box took "much longer" to grab than lighter ones. One participant stated he was initially unable to identify the cause of his sense of weight before noticing that the trigger "was harder to press". However, he stated that he stopped being consciously aware of the change after some time since he felt so immersed in the virtual world.

The remaining two-thirds of participants reported vibrations only, visual input or sound as the basis for their decision process, or they were unable to tell. Vibrations only were mentioned by six participants who further described different intensities, moments of appearance and different ways of how vibrations ended. One participant stated he focused on the vibration and perceived a sound when he experienced more vibrations, despite the noise-cancelling headphones with neutral music. One participant who was unable to identify the reason for his selection process still described the boxes as "much heavier" and "much lighter".

During the interview, participants were also asked if they had spontaneous associations for the content of the boxes during the task. Nine did not, but twelve did. For them, light boxes felt empty and were associated with feathers while heavy boxes felt solid and as if they were filled with sand, stones, gravel, brick or a book. One participant stated he imagined the boxes empty, but made from different materials.

7 DISCUSSION AND FUTURE WORK

The evaluation of Triggermuscle showed that the revised spring mechanism with a larger resistance range improved the perception of adaptive trigger resistance compared to the pilot study using the prototype system. However, it also showed some limitations of the current spring mechanism to simulate weight over variable resistance. In this section, we examine our results by identifying the advantages and limitations of our design from a perceptual and a hardware perspective and laying out potential directions for improvement in hardware design and follow-up studies.

7.1 RQ1—Do Different Trigger Resistances Influence the Haptic Perception of Different Virtual Weights in VR?

The interview responses of our main study revealed seven participants—one third of the sample—self-reporting the change in trigger resistance and discriminating different virtual weights according to the intensity of the resistance. These reports about the weight experience demonstrate that higher resistances were associated with heavier virtual weights, smaller resistances with lighter virtual weights. While this applies only to a limited number of our participants, it presents early indications that different trigger resistances can provide substitutional haptic weight cues at the index finger and, therefore, induce a sense of lighter and heavier virtual objects. However, at the current time, our results remain inconclusive and we cannot fully confirm our assumptions as a considerable number of users were also unable to effectively sense, interpret and associate the cues. This raises the question of why some participants experienced virtual weight using Triggermuscle while others did not.

The quantitative data of our main study highlights the users' ability to discriminate resistances for 13 data sets. This suggests that 3/5 of the participants were able to successfully differentiate between the resistance intensities rendered by Triggermuscle. The data sets consist of four participants with moderate success rates qualifying for the JND calculation, seven participants almost always identifying the heavier box and two additional participants showing an average success rate of 80 and 84%. Surprisingly, this number of data sets exceeds the number of self-reports about the change in trigger resistance by six. This discrepancy may indicate that some participants registered different resistances, but were not consciously aware of it.

7.1.1 Visual Dominance

We believe that the often reported visual dominance in human perception (Posner et al., 1976; Hecht and Reiner, 2009) might have caused the differences in the individual haptic sensing of the trigger resistance and shifted the focus for some participants away from the haptic sensation at their index finger towards the visually identical-looking boxes. Our assumption is supported by the results of our pilot study in which more participants noticed the change in trigger resistance after we simplified the visual input in response to the observations during the preliminary testing. The modification removed the visual sensation of identical-looking boxes which might have previously dominated the perception of the varying haptic feedback. In addition, the visual dominance has been previously reported in VR in the context of pseudo-haptics to overwrite haptic cues, e.g., in haptic retargeting (Azmandian et al., 2016; Zenner et al., 2021), shape rendering (Ban et al., 2012) or the perception of different surface textures in VR (Choi et al., 2018; Degraen et al., 2019). While these approaches purposely made use of this effect, it might have disrupted the haptic perception of the trigger resistance in case of Triggermuscle. We intend to follow this up in future work by exploring the ability to discriminate the trigger resistance in a non-VR setting where participants are blindfolded. This might focus participants' attention on the haptic sense and, further, remove the aspect of weight association. Since this setup would be similar to typical studies investigating the discrimination of stiffness, we would expect the results to be in line with the reported weber fractions from such studies. Nonetheless, we would also like to explore the pseudo-haptics approach through meaningfully combining the trigger resistance with virtual objects that visually indicate different weights. Comparing the weight perception based on only the adaptive resistance, only visual input and a combination of both could identify if visual input could also improve the ability to discriminate between resistances and facilitates an association with virtual weight in VR.

7.1.2 Subtle Vibrations

To better understand participants' discrimination mechanisms and to detect the causes of individual differences, we cross-referenced our qualitative and quantitative results for each participant by comparing the reported decision factors with the psychometric functions. This helped to identify the environmental vibration of the controller's handle, which occurred as an unintended side effect of the servo's adjustment, as a potential haptic confounding factor. Psychometric functions indicating an influence of the resistance on the perception system belonged to interview statements reporting either only the altered trigger resistance, or only vibration, or a combination of both. More precisely, participants who stated that their decision relied entirely on trigger resistance exhibited the expected decrease in their discrimination ability for smaller resistances. While these cases emphasise a relationship between the level of trigger resistance and the perceived weight, those participants who performed best stated both resistance change and vibration, but also vibration on its own as decisive factors. This reveals that vibrations, as an additionally perceived haptic cue, might have interfered with the ability to sense different resistances and their association with weight difference. This assumption, however, is challenged by our data. Five of our participants also stated that they based their decisions on the sensed vibrations while producing poor success rates closely around the guess rate. Due to these two extremes, we cannot make a clear assumption about the influence of vibrations, positive or negative, on the ability to discriminate variable resistances rendered by Triggermuscle and on the association with virtual weight. While previous work has intentionally utilised vibrotactile feedback to render contact forces in addition to asymmetric skin deformation for weight sensation, it has also identified vibration amplitude as a possible confounding factor with an unknown effect on the perception of skin deformation (Choi et al., 2017).

One possible explanation for the differences in differentiating resistances could be the weber fraction of vibrotactile frequency, which ranges widely from 3 to 30% (Jones and Tan, 2013). Some participants might have been, therefore, more receptive to vibrations than others, causing different degrees of distraction away from the change in trigger resistance. This assumption is supported by participants describing vibrations to different extents: One participant reported that he did not notice any vibrations, while others described them as a side effect and yet others focused on them as the main indicator of virtual weight. Since participants were not informed about the adaptive trigger, this might have consolidated a focus on vibration. Further, this possible shift in attention to vibration could also have been promoted by servo adjustment (i.e., the occurrence of vibration). To prevent adjustment during grasping in the main study, the servo changed its angle at the beginning of each trial and before grasping the second box. However, observations during the task showed that some participants released the trigger very slowly and carefully when placing the box. In the case of the first box, the servo's adjustment for the second box was then provoked when participants were still focused on the previous box. In addition, the very first servo adjustment in the initial trial appeared even before participants pulled the trigger for the first time, hence before they experienced any resistance. Nonetheless, the occurrence of vibration does not follow a clear pattern. Vibrations occur not only as soon as the servo registers a pulling force inside the mechanism (e.g., when resistance exceeds the standard value), but also when the trigger is pulled during standard value configuration. Additionally, in this state, a reverberation sometimes occurs when the trigger is released, meaning the servo is active, i.e., causing subtle vibrations. Vibrations as a distraction in haptic devices were previously described as "one of the most noticeable disturbances in a force reflecting device" (Tan et al., 1994). To quantify participants' exposure to the reported subtle vibrations, we took measurements using the digital vibration meter no. 480 600 from VOGEL GERMANY (VOGEL GERMANY, 2021). While all servo adjustments from the main study were tested, only the switch between the minimum and maximum resistances created measurable vibrations. However, these vibrations are below the perceivable range. Importantly, this effect occurs mainly only during the brief period of servo adjustment and does not mature into a continuous vibration.

In future studies, we would like to clarify these possible limitations concerning vibrations. To approach a consistent sensing of the varying resistance, we would like to investigate the role of attention by informing users about the adaptive trigger and by additionally instructing them to ignore the current mechanism's side effect. In addition, shifting the focus away from vibration and to the trigger's intended feedback could be further facilitated by implementing a short adjustment phase. This would ensure that the servo's angle is not configured while participants are close to a virtual object.

7.1.3 Hardware

To address this matter in future hardware design and decrease users' exposure to unintended vibrations, other types of actuation technologies could be tested, such as a micro linear actuator for controlling the spring's length or exploring magnetic repulsion forces which are used in magnetic force-feedback joysticks. This could, additionally, increase the stimulus range and emphasise the trigger resistance as the weight cue. Apart from that, vibrations could also be intentionally used to enhance the experience of virtual weight. A perspective on this matter is demonstrated by previous work which explored vibratory stimulation and patterns for weight perception during the interaction with a vision-tactile-force display (Mizuno et al., 2013). As a movable weight shifts along the display's back towards one of the display's handles, the perceived weight of that handle seems heavier when strong vibrations are rendered. Vibrations were also observed to enhance other virtual object properties such as virtual stiffness in combination with visual information (Maereg et al., 2017). To better understand the possibility to enhance Triggermuscle's weight perception meaningfully with vibrations, further research is necessary.

To achieve a simple weight rendering technique that could potentially be integrated into consumer VR controllers, our approach took only into account the trigger and its level of resistance. While participants were, therefore, presented with different resistive forces at their fingertip, Triggermuscle did not render skin stretch as an additional weight cue. This stimulus is often provided in other haptic devices also focusing on rendering haptic feedback at users' fingers, such as (Minamizawa et al., 2007a; Scheggi et al., 2010; Kurita et al., 2011; Suchoski et al., 2018). While it is common for approaches of sensory substitution to not address the receptors that are addressed in reality, this limitation could have contributed to the different reactions towards the same haptic feedback of Triggermuscle. As the previous finger-worn haptic interfaces for skin stretch use belts or plates stretching and pressing the finger pad, users passively experience the skin modulation in mid-air without physical counter forces to the hand. Since, in contrast, users of the handheld Triggermuscle actively apply pressure on their skin by pulling the trigger and experiencing the counter force of the held device, an additional integration of skin stretch rendering at the trigger might enhance the weight sensation of Triggermuscle. While previous research has shown to successfully render haptic feedback only to users' index fingers for object properties (Benko et al., 2016; Choi et al., 2018; Whitmire et al., 2018; Sinclair et al., 2019), not receiving a sensation to the full hand during grasping might have further impacted the weight sensation of Triggermuscle.

7.2 RQ2—How can the Intensity of the Trigger be Quantified and Mapped to Convey Distinguishable Virtual Weights?

Our findings show that some participants easily detected smaller changes in the resistance, thus enabling them to haptically and precisely render lighter objects, while other participants did not notice any influence of the intensity level even with large differences. Sensitivity differences towards resistance might have been affected by the previously discussed limitations. Nonetheless, self-reports on absolute weight and on the respective box content associations revealed comparable statements for lower and higher resistances. While these reported impressions indicated a possible influence of the visual appearance of the virtual boxes used in the experimental task, they suggest the possibility of using visual input to map the level of resistance and the perceived absolute virtual weight. Future investigating adaptive trigger resistance as a weight metaphor could determine whether haptic feedback could be used to convey a large range of different relative virtual weights. This could be achieved by creating a visual weight context through visually rendering objects of a similar weight class in the VE and through testing the same resistances in different visual weight classes. To advance the experience of virtual weight during interaction, auditory cues could also be presented to convey weight information. For example, a hollow sound could convey an empty, light object whereas a dull sound could convey a filled, heavy object. This arrangement could also be applied to objects of different materials due to different weights.

In summary, designing triggers with adaptive resistance could be one way of equipping VR controllers with enriched haptic feedback in the future. Since the trigger is a commonly used button that can be found in other VR controllers such as Oculus Touch or game controllers, actuated triggers could be integrated into a various interaction devices, also beyond the VR domain. To achieve the intended effect, Triggermuscle's spring mechanism could be tailored to other controller shapes. Alternatively, actuation could be modified with various technical approaches including motors, springs, magnets, and gears, to flexibly fit a wide array of form factors. Particular attention should be paid to the discussed limitations in order to achieve a consistent sensing of the haptic feedback for all users. These limitations should be addressed through hardware design by minimising unintended vibrations. Our findings also indicate possible cross-modal effects, which need to be considered when designing adaptive trigger resistance for VR. One possible limitation of our study is that we asked participants about the object's weight, as this may have biased their association with the provided feedback. Although, participants were not informed of the type of haptic feedback provided, one third linked the stimulus to weight perception. Based on user reports, we further assume that some participants had a real sense of weight as they spontaneously imagined light boxes being filled with feathers, and the heavy ones with sand, stones or gravel.

In the future, we are also interested in expanding the scope of applications for adaptive trigger resistance in VR beyond weight perception by exploring visual cues for other physical properties. As studies (Johansson and Westling, 1984; Flanagan et al., 1995; Ellis and Lederman, 1999) have documented that different material properties impact the level of applied grip forces when lifting an object and its perceived weight, this effect could be potentially be used to further enhance weight rendering in VR and additionally convey different materials of the lifted object. Adapting the trigger resistance then not only to the object's weight but also to its surface roughness or smoothness could account for the naturally performed grip force modulation as reported in the studies and contribute to a more realistic weight experience. Further, we are interested in whether the trigger resistance can substitute haptic cues for virtual stiffness and if different grasp animations can further convey various surface tensions of different materials. We also imagine an integration of resistive forces beyond the trigger in other haptic devices, such as the haptic VR controller Haptic Revolver (Whitmire et al., 2018) to simulate surface stiffness. Users of this device are presented with textured wheels at their index finger to haptically experience shear forces and textures. By adding resistive forces to the textured wheel, varying surface stiffness could be rendered when the textured wheel is pressed down. Additionally, introducing an active modulation of the resistance during the pull of the trigger could render a modulation of surface tension, e.g., depending on how much a deformable material is squeezed. The resistance modulation could also convey weight shifts, e.g., as the liquid inside a cup moves around when balancing it or when an object is accelerated in order to be thrown. The brief increase or decrease of trigger resistance could furthermore be applied for haptic feedback of operating virtual UI elements.

8 CONCLUSION

We explored adaptive trigger resistance as a novel approach for weight perception in VR. The adaptive trigger inside our haptic VR controller *Triggermuscle* modifies the level of resistance according to the weight of a grabbed virtual object. Users, therefore, need to scale their index finger force to grab objects with different virtual weights. We presented the design and implementation of the prototype's initial and Triggermuscle's revised spring mechanism, which adjusts the trigger resistance dynamically and continuously. As proof of concept, our systems are built into the casing of an HTC Vive controller. The mechanisms, however, can easily be adapted to various of form factors of different interaction devices. In two user studies, we evaluated both technical implementations and explored the positive as well as the negative effects of actuated adaptive trigger resistance on users' discrimination ability and associations with virtual object weight in VR. Our findings show large differences between participants. While Triggermuscle's adaptive trigger enabled 3 out of 5 participants to sense and discriminate different levels of resistance, only 1 out of 3 participants self-reported an association with weight. In the successful weight illusion, lower resistances were associated with lighter objects and higher resistances with heavier objects. While these findings reveal early indications for using adaptive trigger resistance to simulate virtual weight in VR, they reveal limitations regarding the sensing of variable resistance, perceptual mechanisms and hardware design. We have discussed and wish to address these limitations in future studies. Our findings provide a first important step towards using adaptive trigger resistance in VR. We hope that this work motivates further research on transforming established input elements into input-output components, so as to enhance the haptic experience of handheld VR controllers.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

CS was the primary author and developed both systems, conducted user studies, data analysis, literature review, manuscript writing and manuscript review. MB contributed to the discussion of the research idea as well as the literature review, manuscript writing and manuscript review. EK contributed to the discussion of the research idea, provided feedback on the prototypes, and took part in manuscript writing and manuscript review. JS contributed to the discussion of the research idea, provided feedback on the prototypes, took part in the manuscript writing and manuscript review and supervised the project. All authors contributed to the article and approved the submitted version.

FUNDING

This work was partially funded by a Lichtenberg Professorship of the Volkswagen Foundation as well as the BMBF project InviDas (16SV853). In addition, the Klaus Tschira Foundation supported a co-author with a PhD scholarship.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frvir.2021.754511/full#supplementary-material

REFERENCES

Azmandian, M., Hancock, M., Benko, H., Ofek, E., and Wilson, A. D. (2016). "Haptic Retargeting: Dynamic Repurposing of Passive Haptics for Enhanced Virtual Reality Experiences," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA: Association for Computing Machinery), 1968–1979. CHI '16. doi:10.1145/2858036.2858226

Ban, Y., Kajinami, T., Narumi, T., Tanikawa, T., and Hirose, M. (2012). "Modifying an Identified Curved Surface Shape Using Pseudo-haptic Effect," in Proceeding of the 2012 IEEE Haptics Symposium (HAPTICS), Vancouver, BC, Canada, 4-7 March 2012 (Vancouver, BC: IEEE), 211–216. doi:10.1109/HAPTIC.2012.6183793

Ben-Tzvi, P., and Ma, Z. (2015). Sensing and Force-Feedback Exoskeleton (SAFE) Robotic Glove. *IEEE Trans. Neural Syst. Rehabil. Eng.* 23, 992–1002. doi:10.1109/TNSRE.2014.2378171

Benko, H., Holz, C., Sinclair, M., and Ofek, E. (2016). "NormalTouch and TextureTouch: High-fidelity 3D Haptic Shape Rendering on Handheld Virtual Reality Controllers," in *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (New York, NY, USAUIST '16: Association for Computing Machinery), 717–728. doi:10.1145/2984511.2984526

Bonfert, M., Porzel, R., and Malaka, R. (2019). "Get a Grip! Introducing Variable Grip for Controller-Based VR Systems," in Proceeding of the 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Osaka, Japan, 23-27 March 2019 (IEEE), 604–612. doi:10.1109/VR.2019.8797824

Bouzit, M., Burdea, G., Popescu, G., and Boian, R. (2002). The Rutgers Master II-New Design Force-Feedback Glove. *Ieee/asme Trans. Mechatron.* 7, 256–263. doi:10.1109/TMECH.2002.1011262

Brodie, E. E., and Ross, H. E. (1984). Sensorimotor Mechanisms in Weight Discrimination. *Perception & Psychophysics* 36, 477–481. doi:10.3758/BF03207502

Burdea, G., Zhuang, J., Roskos, E., Silver, D., and Langrana, N. (1992). A Portable Dextrous Master with Force Feedback. *Presence: Teleoperators & Virtual Environments* 1, 18–28. doi:10.1162/pres.1992.1.1.18

Cheng, C.-H., Chang, C.-C., Chen, Y.-H., Lin, Y.-L., Huang, J.-Y., Han, P.-H., et al. (2018). "GravityCup: A Liquid-Based Haptics for Simulating Dynamic Weight in Virtual Reality," in *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology* (New York, NY, USA: Association for Computing Machinery), 1–2. VRST '18. doi:10.1145/3281505.3281569

Choi, I., Culbertson, H., Miller, M. R., Olwal, A., and Follmer, S. (2017). "Grabity: A Wearable Haptic Interface for Simulating Weight and Grasping in Virtual Reality," in *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA: Association for Computing Machinery), 119–130. UIST '17. doi:10.1145/3126594.3126599

Choi, I., Ofek, E., Benko, H., Sinclair, M., and Holz, C. (2018). "CLAW: A Multifunctional Handheld Haptic Controller for Grasping, Touching, and Triggering in Virtual Reality," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA: Association for Computing Machinery). CHI '18, 1–13. doi:10.1145/3173574.3174228

Degraen, D., Zenner, A., and Krüger, A. (2019). "Enhancing Texture Perception in Virtual Reality Using 3D-Printed Hair Structures," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA: Association for Computing Machinery), 1–12. CHI '19. doi:10.1145/3290605.3300479

Dominjon, L., Lecuyer, A., Burkhardt, J., Richard, P., and Richir, S. (2005). "Influence of Control/Display Ratio on the Perception of Mass of Manipulated Objects in Virtual Environments," in Proceedings - IEEE Virtual Reality, Bonn, Germany, 12-16 March 2005 (IEEE). doi:10.1109/vr.2005.1492749

Ellis, R. R., and Lederman, S. J. (1999). The Material-Weight Illusion Revisited. *Perception & Psychophysics* 61, 1564–1576. doi:10.3758/BF03213118

Ellis, R. R., and Lederman, S. J. (1993). The Role of Haptic versus Visual Volume Cues in the Size-Weight Illusion. *Perception & Psychophysics* 53, 315–324. doi:10.3758/BF03205186

Fang, C., Zhang, Y., Dworman, M., and Harrison, C. (2020). "Wireality: Enabling Complex Tangible Geometries in Virtual Reality with Worn Multi-String Haptics," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA: Association for Computing Machinery). CHI '20, 1–10. doi:10.1145/3313831.3376470

Flanagan, J. R., Wing, A. M., Allison, S., and Spenceley, A. (1995). Effects of Surface Texture on Weight Perception when Lifting Objects with a Precision Grip. *Perception & Psychophysics* 57, 282–290. doi:10.3758/BF03213054

Gabardi, M., Solazzi, M., Leonardis, D., and Frisoli, A. (2016). "A New Wearable Fingertip Haptic Interface for the Rendering of Virtual Shapes and Surface Features," in Proceeding of the 2016 IEEE Haptics Symposium, HAPTICS, Philadelphia, PA, USA, 8-11 April 2016 (IEEE), 140–146. doi:10.1109/haptics.2016.7463168

Giannopoulos, E., Pomes, A., and Slater, M. (2012). Touching the Void: Exploring Virtual Objects through a Vibrotactile Glove. *Ijvr* 11, 19–24. doi:10.20870/IJVR.2012.11.3.2847

Girard, A., Marchal, M., Gosselin, F., Chabrier, A., Louveau, F., and Lécuyer, A. (2016). Haptip: Displaying Haptic Shear Forces at the Fingertips for Multi-finger Interaction in Virtual Environments. *Front. ICT* 3, 718. doi:10.3389/fict.2016.00006

Gordon, A. M., Forssberg, H., Johansson, R. S., and Westling, G. (1991). Visual Size Cues in the Programming of Manipulative Forces during Precision Grip. *Exp. Brain Res.* 83, 477–482. doi:10.1007/BF00229824

Gu, X., Zhang, Y., Sun, W., Bian, Y., Zhou, D., and Kristensson, P. O. (2016). "Dexmo: An Inexpensive and Lightweight Mechanical Exoskeleton for Motion Capture and Force Feedback in VR," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA: Association for Computing Machinery), 1991–1995. CHI '16. doi:10.1145/2858036.2858487

Guilford, J. P. (1955). *Psychometric Methods*. 2nd edn. New York: McGraw-Hill.

HaptX (2020). *Haptx Gloves*. Available at: https://haptx.com/ (Accessed June 15, 2020).

Hecht, D., and Reiner, M. (2009). Sensory Dominance in Combinations of Audio, Visual and Haptic Stimuli. *Exp. Brain Res.* 193, 307–314. doi:10.1007/s00221-008-1626-z

Heo, S., Chung, C., Lee, G., and Wigdor, D. (2018). "Thor's Hammer: An Ungrounded Force Feedback Device Utilizing Propeller-Induced Propulsive Force," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA: Association for Computing Machinery), 1–11.

Hirose, M., Hirota, K., Ogi, T., Yano, H., Kakehi, N., Saito, M., et al. (2001). "Hapticgear: the Development of a Wearable Force Display System for Immersive Projection Displays," in Proceedings IEEE Virtual Reality 2001, Yokohama, Japan, 13-17 March 2001 (IEEE), 123–129. doi:10.1109/VR.2001.913778

Ishii, M., and Sato, M. (1994). A 3D Spatial Interface Device Using Tensed Strings. *Presence: Teleoperators & Virtual Environments* 3, 81–86. doi:10.1162/pres.1994.3.1.81

Je, S., Kim, M. J., Lee, W., Lee, B., Yang, X.-D., Lopes, P., et al. (2019). "Aero-plane: A Handheld Force-Feedback Device that Renders Weight Motion Illusion on a Virtual 2d Plane," in *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA: Association for Computing Machinery), 763–775. UIST '19. doi:10.1145/3332165.3347926

Jenmalm, P., and Johansson, R. S. (1997). Visual and Somatosensory Information about Object Shape Control Manipulative Fingertip Forces. *J. Neurosci.* 17, 4486–4499. doi:10.1523/jneurosci.17-11-04486.1997

Johansson, R. S., and Westling, G. (1988). Coordinated Isometric Muscle Commands Adequately and Erroneously Programmed for the Weight during Lifting Task with Precision Grip. *Exp. Brain Res.* 71, 59–71. doi:10.1007/BF00247522

Johansson, R. S., and Westling, G. (1984). Roles of Glabrous Skin Receptors and Sensorimotor Memory in Automatic Control of Precision Grip when Lifting Rougher or More Slippery Objects. *Exp. Brain Res.* 56, 550–564. doi:10.1007/BF00237997

Jones, L. A., and Tan, H. Z. (2013). Application of Psychophysical Techniques to Haptic Research. *IEEE Trans. Haptics* 6, 268–284. doi:10.1109/TOH.2012.74

Kaczmarek, K. A., Webster, J. G., Bach-y-Rita, P., and Tompkins, W. J. (1991). Electrotactile and Vibrotactile Displays for Sensory Substitution Systems. *IEEE Trans. Biomed. Eng.* 38, 1–16. doi:10.1109/10.682010.1109/10.68204

Kurita, Y., Yonezawa, S., Ikeda, A., and Ogasawara, T. (2011). "Weight and Friction Display Device by Controlling the Slip Condition of a Fingertip," in Proceeding of the 2011 IEEE International Conference on Intelligent Robots and Systems, San Francisco, CA, USA, 25-30 Sept. 2011 (IEEE), 2127–2132. doi:10.1109/iros.2011.6094613

Lee, J., Sinclair, M., Gonzalez-Franco, M., Ofek, E., and Holz, C. (2019). "TORC: A Virtual Reality Controller for In-Hand High-Dexterity Finger Interaction," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA: Association for Computing Machinery), 1–13. CHI '19. doi:10.1145/3290605.3300301

Leonardis, D., Solazzi, M., Bortone, I., and Frisoli, A. (2015). "A Wearable Fingertip Haptic Device with 3 Dof Asymmetric 3-rsr Kinematics," in Proceeding of the 2015 IEEE World Haptics Conference (WHC), Evanston, IL, USA, 22-26 June 2015 (IEEE), 388–393. doi:10.1109/WHC.2015.7177743

Liu, Y., Hashimoto, T., Yoshida, S., Narumi, T., Tanikawa, T., and Hirose, M. (2019). "ShapeSense: A 2D Shape Rendering VR Device with Moving Surfaces that Controls Mass Properties and Air Resistance," in *ACM SIGGRAPH 2019 Emerging Technologies* (New York, NY, USA: Association for Computing Machinery), 1–2. SIGGRAPH '19. doi:10.1145/3305367.3327991

Loomis, J. M., and Lederman, S. J. (1986). "Tactual Perception," in *Handbook of Perception and Human Performances* (New York, NY: John Wiley & Sons), 2, 1–41.

Lopes, P., You, S., Cheng, L.-P., Marwecki, S., and Baudisch, P. (2017). "Providing Haptics to Walls & Heavy Objects in Virtual Reality by Means of Electrical Muscle Stimulation," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA: Association for Computing Machinery), 1471–1482. CHI '17. doi:10.1145/3025453.3025600

Maereg, A. T., Nagar, A., Reid, D., and Secco, E. L. (2017). Wearable Vibrotactile Haptic Device for Stiffness Discrimination during Virtual Interactions. *Front. Robot. AI* 4, 42. doi:10.3389/frobt.2017.00042

Maiero, J., Eibich, D., Kruijff, E., Hinkenjann, A., Stuerzlinger, W., Benko, H., et al. (2019). Back-of-Device Force Feedback Improves Touchscreen Interaction for Mobile Devices. *IEEE Trans. Haptics* 12, 483–496. doi:10.1109/toh.2019.2911519

Maisto, M., Pacchierotti, C., Chinello, F., Salvietti, G., De Luca, A., and Prattichizzo, D. (2017). Evaluation of Wearable Haptic Systems for the Fingers in Augmented Reality Applications. *IEEE Trans. Haptics* 10, 511–522. doi:10.1109/TOH.2017.2691328

Manus (2020). *Prime II Mocap Gloves*. Available at: https://www.manus-vr.com/mocapgloves (Accessed June 17, 2020).

Marquardt, A., Kruijff, E., Trepkowski, C., Maiero, J., Schwandt, A., Hinkenjann, A., et al. (2018). "Audio-Tactile Proximity Feedback for Enhancing 3D Manipulation," in *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology* (New York, NY, USA: Association for Computing Machinery), 1–10. VRST '18. doi:10.1145/3281505.3281525

Martínez, J., García, A., Oliver, M., Molina, J. P., and González, P. (2016). Identifying Virtual 3d Geometric Shapes with a Vibrotactile Glove. *IEEE Computer Graphics Appl.* 36, 42–51. doi:10.1109/MCG.2014.81

Massie, T. H., and Salisbury, J. K. (1994). "The Phantom Haptic Interface: A Device for Probing Virtual Objects," in *Proceedings of the ASME Winter Annual Meeting, Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems* (Chicago), 295–300.

McCloskey, D. I. (1974). Muscular and Cutaneous Mechanisms in the Estimation of the Weights of Grasped Objects. *Neuropsychologia* 12, 513–520. doi:10.1016/0028-3932(74)90081-5

Microsoft (2021). *Xbox Elite Wireless Controller*. Available at: https://www.xbox.com/en-US/accessories/controllers/elite-wireless-controller (Accessed June 3, 2021).

Minamizawa, K., Fukamachi, S., Kajimoto, H., Kawakami, N., and Tachi, S. (2007a). "Gravity Grabber: Wearable Haptic Display to Present Virtual Mass Sensation," in *ACM SIGGRAPH 2007 Emerging Technologies* (New York, NY, USA: Association for Computing Machinery), 3–6. SIGGRAPH '07. doi:10.1145/1278280.1278289

Minamizawa, K., Kajimoto, H., Kawakami, N., and Tachi, S. (2007b). "A Wearable Haptic Display to Present the Gravity Sensation - Preliminary Observations and Device Design," in Proceeding of the 2013 Second Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (WHC'07), Tsukuba, Japan, 22-24 March 2007 (IEEE), 133–138. doi:10.1109/WHC.2007.15

Mizuno, T., Maeda, J., and Kume, Y. (2013). "Weight Sensation Affected by Vibrotactile Stimulation with a Handheld Vision-Tactile-Force Display Device," in Proceeding of the 2013 10th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, Krabi, Thailand, 15-17 May 2013 (IEEE), 1–6. doi:10.1109/ECTICon.2013.6559590

Nisar, S., Martinez, M. O., Endo, T., Matsuno, F., and Okamura, A. M. (2019). Effects of Different Hand-Grounding Locations on Haptic Performance with a Wearable Kinesthetic Haptic Device. *IEEE Robot. Autom. Lett.* 4, 351–358. doi:10.1109/lra.2018.2890198

Pacchierotti, C., Salvietti, G., Hussain, I., Meli, L., and Prattichizzo, D. (2016). "The hRing: A Wearable Haptic Device to Avoid Occlusions in Hand Tracking," in Proceeding of the 2016 IEEE Haptics Symposium (HAPTICS), Philadelphia, PA, USA, 8-11 April 2016 (IEEE), 134–139. doi:10.1109/haptics.2016.7463167

Posner, M. I., Nissen, M. J., and Klein, R. M. (1976). Visual Dominance: An Information-Processing Account of its Origins and Significance. *Psychol. Rev.* 83, 157–171. doi:10.1037/0033-295X.83.2.157

Provancher, W. (2014). Creating Greater VR Immersion by Emulating Force Feedback with Ungrounded Tactile Feedback. *IQT Quarterly* 6, 18–21.

Rietzler, M., Geiselhart, F., Gugenheimer, J., and Rukzio, E. (2018). "Breaking the Tracking: Enabling Weight Perception using Perceivable Tracking Offsets," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA: Association for Computing Machinery), 1–12. CHI '18. doi:10.1145/3173574.3173702

Ryu, N., Lee, W., Kim, M. J., and Bianchi, A. (2020). "ElaStick: A Handheld Variable Stiffness Display for Rendering Dynamic Haptic Response of Flexible Object," in *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (New York, NY: Association for Computing Machinery). UIST '20. doi:10.1145/3379337.3415862

Samad, M., Gatti, E., Hermes, A., Benko, H., and Parise, C. (2019). "Pseudo-Haptic Weight: Changing the Perceived Weight of Virtual Objects By Manipulating Control-Display Ratio," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA: Association for Computing Machinery). CHI '19, 1–13. doi:10.1145/3290605.3300550

Sato, M. (2002). Development of String-Based Force Display: Spidar. *8th International Conference on Virtual Systems and Multimedia*. 1034–1039. Citeseer.

Scheggi, S., Salvietti, G., and Prattichizzo, D. (2010). "Shape and Weight Rendering for Haptic Augmented Reality," in Proceedings - IEEE International Workshop on Robot and Human Interactive Communication, Viareggio, Italy, 13-15 Sept. 2010 (IEEE), 44–49. doi:10.1109/roman.2010.5598632

Schorr, S. B., and Okamura, A. M. (2017). "Fingertip Tactile Devices for Virtual Object Manipulation and Exploration," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (New York, NY: Association for Computing Machinery), 3115–3119. CHI '17. doi:10.1145/3025453.3025744

Schütt, H. H., Harmeling, S., Macke, J. H., and Wichmann, F. A. (2016). Painfree and Accurate Bayesian Estimation of Psychometric Functions for (Potentially) Overdispersed Data. *Vis. Res.* 122, 105–123. doi:10.1016/j.visres.2016.02.002

Schütt, H. H. (2019). *Psignifit*. Available at: https://github.com/wichmann-lab/psignifit/wiki (Accessed June 16, 2020).

Shigeyama, J., Hashimoto, T., Yoshida, S., Narumi, T., Tanikawa, T., and Hirose, M. (2019). "Transcalibur: A Weight Shifting Virtual Reality Controller for 2D Shape Rendering based on Computational Perception Model," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA: Association for Computing Machinery), 1–11. CHI '19. doi:10.1145/3290605.3300241

Simpson, W. A. (1988). The Method of Constant Stimuli Is Efficient. *Perception & Psychophysics* 44, 433–436. doi:10.3758/bf03210427

Sinclair, M., Ofek, E., Gonzalez-Franco, M., and Holz, C. (2019). "CapstanCrunch: A Haptic VR Controller with User-supplied Force Feedback," in *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA: Association for Computing Machinery), 815–829. UIST '19. doi:10.1145/3332165.3347891

Solazzi, M., Frisoli, A., Salsedo, F., and Bergamasco, M. (2007). "A Fingertip Haptic Display for Improving Local Perception of Shape Cues," in Proceeding of the Second Joint EuroHaptics Conference and Symposium on Haptic Interfaces for

Virtual Environment and Teleoperator Systems (WHC'07), Tsukuba, Japan, 22-24 March 2007 (IEEE), 409–414. doi:10.1109/WHC.2007.3
Sony (2020). *Dualsense Wireless Controller*. Available at: https://www.playstation.com/en-gb/accessories/dualsense-wireless-controller/ (Accessed September 4, 2020).
Stellmacher, C. (2021). "Haptic-Enabled Buttons through Adaptive Trigger Resistance," in Proceeding of the 2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), Lisbon, Portugal, 27 March-1 April 2021 (IEEE), 201–204. doi:10.1109/VRW52623.2021.00044
Strasnick, E., Holz, C., Ofek, E., Sinclair, M., and Benko, H. (2018). "Haptic Links: Bimanual Haptics for Virtual Reality Using Variable Stiffness Actuation," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA: Association for Computing Machinery), 1–12.
Suchoski, J. M., Martinez, S., and Okamura, A. M. (2018). "Scaling Inertial Forces to Alter Weight Perception in Virtual Reality," in Proceeding of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, QLD, Australia, 21-25 May 2018 (IEEE), 484–489. doi:10.1109/icra.2018.8462874
Sun, Y., Yoshida, S., Narumi, T., and Hirose, M. (2019). "PaCaPa: A Handheld VR Device for Rendering Size, Shape, and Stiffness of Virtual Objects in Tool-based Interactions," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA: Association for Computing Machinery), 1–12. CHI '19. doi:10.1145/3290605.3300682
3D Systems (2021). *Phantom Premium*. Available at: https://www.3dsystems.com/haptics-devices/3d-systems-phantom-premium (Accessed June 15, 2020).
Tan, H. Z., Srinivasan, M. A., Eberman, B., and Cheng, B. (1994). "Human Factors for the Design of Force-Reflecting Haptic," in *Dynamics Systems and Control* (American Society of Mechanical Engineers (ASME)), 55, 353–359.
Tsai, H.-R., Rekimoto, J., and Chen, B.-Y. (2019). "ElasticVR: Providing Multilevel Continuously-Changing Resistive Force and Instant Impact Using Elasticity for VR," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (New York, NY: Association for Computing Machinery), 1–10. CHI'19. doi:10.1145/3290605.3300450
Van Polanen, V., and Davare, M. (2015). Sensorimotor Memory Biases Weight Perception during Object Lifting. *Front. Hum. Neurosci.* 9, 1–11. doi:10.3389/fnhum.2015.00700
VOGEL GERMANY (2021). Electr. Digital Vibration Meter. Available at: https://www.vogel-germany.de/en/produkte-kataloge-prospekte-edition.html.
Westling, G., and Johansson, R. S. (1984). Factors Influencing the Force Control during Precision Grip. *Exp. Brain Res.* 53, 277–284. doi:10.1007/BF00238156
Whitmire, E., Benko, H., Holz, C., Ofek, E., and Sinclair, M. (2018). "Haptic Revolver: Touch, Shear, Texture, and Shape Rendering on a Reconfigurable Virtual Reality Controller," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA: Association for Computing Machinery), 1–12. CHI '18. doi:10.1145/3173574.3173660
Yu, R., and Bowman, D. A. (2020). Pseudo-haptic Display of Mass and Mass Distribution during Object Rotation in Virtual Reality. *IEEE Trans. Vis. Comput. Graphics* 26, 2094–2103. doi:10.1109/TVCG.2020.2973056
Zenner, A., and Krüger, A. (2019). "Drag:on: A Virtual Reality Controller Providing Haptic Feedback Based on Drag and Weight Shift," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA: Association for Computing Machinery), 1–12. CHI '19. doi:10.1145/3290605.3300441
Zenner, A., and Krüger, A. (2017). Shifty: A Weight-Shifting Dynamic Passive Haptic Proxy to Enhance Object Perception in Virtual Reality. *IEEE Trans. Vis. Comput. Graphics* 23, 1285–1294. doi:10.1109/TVCG.2017.2656978
Zenner, A., Ullmann, K., and Krüger, A. (2021). Combining Dynamic Passive Haptics and Haptic Retargeting for Enhanced Haptic Feedback in Virtual Reality. *IEEE Trans. Vis. Comput. Graphics* 27, 2627–2637. doi:10.1109/TVCG.2021.3067777

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Publication P6

Kicking in Virtual Reality: The Influence of Foot Visibility on the Shooting Experience and Accuracy

Michael Bonfert, Stella Lemke, Robert Porzel, and Rainer Malaka

This publication investigates how the visibility of the player's foot influences penalty shooting in VR soccer. In a between-group experiment, we asked 28 players to hit eight targets with a virtual ball. The foot's visual representation improved self-reported body ownership and the accuracy of the shots significantly. Players with invisible foot needed 58% more attempts.

My contribution: I contributed the research question (*conceptualization*) and statistical analysis (*data curation* and *formal analysis*). I contributed the majority of the study design (*methodology*). I wrote and edited the majority of the manuscript (*writing – draft, review, & editing*) and created all the figures and presentation material (*visualization*). A made minor contributions to the prototype development (*software*). I supervised the project (*supervision*).

Published in the *Proceedings of the 2022 IEEE Conference on Virtual Reality and 3D User Interfaces (IEEE VR '22)*

Kicking in Virtual Reality:
The Influence of Foot Visibility on the Shooting Experience and Accuracy

Michael Bonfert Stella Lemke Robert Porzel Rainer Malaka

Digital Media Lab, University of Bremen, Germany*

Figure 1: (left) A side view of the player's visualized foot before kicking the virtual ball. (right) The targets to shoot. The second window from left was just hit and breaks. The fourth is flashing as indicator to be shot next. The player perspective is further away and centered.

Abstract

When playing sports in virtual reality foot interaction is crucial for many disciplines. We investigated how the visibility of the foot influences penalty shooting in soccer. In a between-group experiment, we asked 28 players to hit eight targets with a virtual ball. We measured the performance, task load, presence, ball control, and body ownership of inexperienced to advanced soccer players. In one condition, the players saw a visual representation of their tracked foot which improved the accuracy of the shots significantly. Players with invisible foot needed 58% more attempts. Further, with foot visibility the self-reported body ownership was higher.

Index Terms: Human-centered computing—Virtual reality; Human-centered computing—Empirical studies in HCI;

1 Introduction

Foot-based interaction in virtual reality (VR) has received far less attention in research and industrial applications compared to hand interaction. In reality, however, we use our feet for a variety of purposes beyond locomotion. Especially in the world of sports, kicking objects is a popular recreational activity as well as a professional enterprise. The amply named game of soccer, or *football* in UK English, even punishes hand-based interactions. Players are passing and playing balls almost completely with their feet.

When shooting at or towards a designated target, professional soccer players neither look at the ball nor their foot. Instead, they lock their view onto the target when taking the shot to improve their aiming. In reality, they can feel the response of the ball through haptic feedback while dribbling and shooting the ball. When simulating soccer games in VR, the impact of the virtual ball cannot be perceived on the player's foot, unless some form of force feedback is rendered, e.g., with active actuators in the shoe [25] or ball props [4].

In order not to presuppose any custom hardware, we set out to investigate precise target shooting by kicking a virtual soccer ball in VR using only a Vive Tracker as input device. We present an empirical study that explores the effects of a visual representation of the player's virtual foot on the ball interaction, as corresponding visualizations of virtual hands have been shown to improve the feeling of presence and body ownership [5, 10, 19]. We therefore ask, if the visibility of the virtual foot makes up for the lack of haptic feedback and facilitates more accurate shots or if it makes no difference as players do not look at their foot while kicking, anyway, as they rely on proprioceptive cues.

To assess the influence of the foot visualization, we designed a task inspired from penalty shooting for a between-groups experiment and measured the players' performances in hitting the targets, the perceived task load, the sense of presence, as well as two aspects of embodiment: the sense of agency, operationalized as action and intention with the kicking foot and therefore the effective control over the ball, and the sense of body ownership, as the experience of the virtual foot as one's own [15, 24]. We hypothesize that a visual representation of the player's foot (H1) improves the performance in terms of accuracy and required time, (H2) reduces the task load, (H3) enhances the presence in the virtual environment, (H4) increases the subjective control over the ball, and (H5) enhances the perceived body ownership.

2 Related Work

Interactive games based on various types of sport and physical activities have been on the market for a long time, for example, the Wii Sports games released by Nintendo in 2006. Here the interaction is based on controllers and additional devices that can be used as a tennis racket, golf club, or baseball bat. The rise of VR headsets for the consumer market, such as the Oculus Rift or HTC Vive, paved the way for VR-based sportive games [40] and physiotherapeutic applications [2] in immersive virtual environments. Consequently,

*{bonfert, slemke, porzel, malaka} @uni-bremen.de

This preprint is identical to the published paper. The original is available at https://doi.org/10.1109/VR51125.2022.00092

studies have investigated the difference between virtual and real sports in high-performance environments, e.g., for golfing [12].

Handling virtual objects without any haptic feedback from a real object presents a major challenge for sports in VR, that also applies to playing with a ball. Zindulka et al. compared throwing a physical ball in reality to throwing a virtual ball with a controller in VR yielding in lower precision and accuracy for virtual throws [41]. Sports that are performed with the foot, such as soccer, have not received the same attention as hand interactions. Naturally, manual interactions are more flexible in their operational abilities and application versatility as well as easier to track – with widely available controllers or optically with cameras built into the headset. While foot interaction has a rich history in human-computer interaction [37], literature on foot interaction in VR often considers the feet as abstract input modality for locomotion [38], navigational tasks [6], or game mechanics such as sneaking [7].

Therefore, there is ample space for more research on foot interactions across various domains, applications, and modalities. Besides efforts in exploring aspects of visual perception, there is a growing body of research on haptic foot displays [31, 32], auditory cues from the feet [14, 33], and multimodal systems [34]. Various approaches for camera-based foot interaction with mobile devices were implemented to realize navigational interfaces [36], interactive control with the foot [21], and kicking gestures [11].

Prior work has examined the effects of visualizing body parts in VR with a focus in the literature on the user's hands, e.g., while typing on a keyboard [10, 16]. Additionally, different degrees of realism of hand representations have been compared showing an increase in agency along with the realism [3] and higher presence along with visual match [28]. Pan and Steed explored the effects of foot visualization in a multi-user scenario. In their study, participants are asked to assemble a jigsaw puzzle in a shared virtual environment with either no foot representation, feet floating below the torso, or accurate foot tracking. They found that the visualization with a self-avatar enabled the participants to assemble the puzzle more quickly and had important effects on presence and interaction. Participants with tracked feet moved closer to obstacles, participants with no visual representation usually ignored them [22].

The fundamental influence of a user's virtual body on presence was also demonstrated in a psychophysical experiment by Skarbez et al. [29]. Like in our study, the participants were instructed to play with a ball with their feet. The physical realism of the ball interaction was varied across experiment conditions. Further, the authors varied if the virtual body was visible and if body parts above the hip were movable. The feet, however, were visible and movable in every condition. The results show that the visibility of the virtual body and the ability to control its movements were the most important contributors to plausibility illusion [30] and therefore presence, although the feet that were required for the kicking activity were always visible and controllable.

In a sports context, researchers explored climbing at physical walls while being in VR with the hands and feet tracked and visualized [17, 26, 35]. In a study by Kosmalla et al., participants had to climb four different routes with four different conditions: no visible hands and feet, only hands, only feet, or hands and feet. The results show that experienced climbers did not need a visualization of the hands but that visible feet alone could reduce the number of missteps [17]. This suggests that proprioceptive cues might not be sufficient for assessing the locations of one's feet in VR sports.

3 User Study Design

To evaluate the effects of visualization on foot interaction in VR, we conducted a user study ($N = 28$) with a kicking task inspired from soccer penalty shootouts. The players' objective was to kick a soccer ball from a stationary position and hit eight targets. In the between-groups experiment, we compared the shooting performance,

Figure 2: As shown in this illustration an HTC Vive Tracker 2.0 was attached to the top of the players' right shoe with elastic bands.

task load, presence, ball control, and body ownership between two conditions: kicking a ball with a visible foot (+F) *versus* with an invisible foot (–F). We decided against abstract indicators for the control condition, e.g., a dot at the tip of the foot [16], because we wanted to learn whether the player's proprioceptive impression is enough or if it needs to be supported by a visual impression of the foot position. Further, we decided against a within-subject study design because we expected strong learning effects, especially from players who would first play with visible foot and then transfer this experience to the condition with invisible foot.

The players were supposed to control the direction of the shot precisely, however, we did not expect them to control the height of the shot. Therefore, the targets were 2.40 m tall and only 1.12 m wide which presented a challenge at 11.22 m distance. The targets were the windows along the front of a house. By hitting a window, the breaking glass served as intuitive feedback for success and, furthermore, as a tempting motivational factor.

3.1 Sample

We recruited 28 participants for this experiment (17 male, 11 female) through social media and soccer club channels. The players were between 17 and 59 years of age ($M = 30.1$, $SD = 13.5$). The sample included inexperienced and advanced soccer players covering the range with an even distribution. The group means of the self-assessments from 1 (beginner) to 5 (professional) are $M_{+F} = 3.36$ and $M_{-F} = 3.07$. All participants except one were new to VR. The random distribution between the conditions was balanced in terms of experience in soccer, VR, and computer games, as well as gender and age, with no significant differences revealed in t, U, and χ^2 tests (all with $p \geq .548$).

3.2 Experimental Setup

The prototype was implemented with Unity 2020.1.1f1 and the SteamVR 2.6.1 plugin. The SteamVR Tracking of the HTC Vive is highly accurate, easy to set up, and can be used without considerable constraints in movement. It and has been used in research for tasks ranging from climbing [26] to learning to dance [20]. In our study, the players were provided with an HTC Vive headset, a Vive controller, and a Vive Tracker attached to the right foot with elastic bands as depicted in Figure 2. With the controller, the player could manually reset the ball position. Additionally, it was reset automatically when the ball left a specified playing field or after hitting the target.

When the ball was kicked, upon collision of the shoe model's mesh collider and the ball's sphere collider, an impulse was calculated to enable predictable and powerful shots: the foot's current relative velocity was multiplied with 550 and transferred as impulse

Table 1: The means, standard deviations, group differences, test statistics and significance values for all statistical comparisons between the two conditions visible foot (+F) and invisible foot (–F). Significant differences in these one-sided tests with $p < .05$ are marked with an asterisk.

			Mean	*SD*	Δ	Test statistics	*p*
Performance	Time (min)	+F	6:21	2:14	1:17	$t\,(26) = -1.286$	.105
		–F	7:38	2:58			
	Accuracy (attempts)	+F	52.50	21.27	30.14	$t\,(17.96) = -2.157$	**.022***
		–F	82.64	47.78			
	Rate (attempts/min)	+F	8.45	2.92	2.03	$t\,(26) = -1.674$	.947
		–F	10.48	3.47			
Task Load Index		+F	30.48	7.70	3.03	$U = 81.0$	.224
		–F	33.51	6.57			
Presence	Involvement/Control	+F	5.54	0.59	0.01	$U = 93.5$	.591
		–F	5.53	0.41			
	Naturalness	+F	4.81	1.29	0.10	$U = 109.5$	.306
		–F	4.71	1.04			
	Interface Quality	+F	4.76	1.26	0.34	$U = 83.5$	.756
		–F	5.10	0.88			
Ball Control		+F	3.72	1.39	0.02	$U = 93.0$	.600
		–F	3.74	0.80			
Body Ownership		+F	5.51	1.10	1.07	$U = 136.5$	**.040***
		–F	4.44	1.60			

to the ball's rigidbody. This factor was determined empirically in iterative pre-tests by testing which impulse provided most test users with the impression of realistic shooting physics. The subsequent trajectory was calculated by Unity's physics engine.

The players in the condition with invisible foot could not see any visualization of their foot or indicator of its location. The players with foot representation could see the virtual soccer shoe and sock depicted in Figure 1. Since the foot was only tracked in one place, the ankle joint was not dynamically animated. All players were right-footed, hence, all players shot with their right foot. The left foot was neither tracked nor virtually displayed.

The virtual scene was situated outdoors. The player stood on a lawn looking at a house with target windows along its front. The spacious playing field was enclosed by trees and fences. The windows had to be hit one after the other as indicated with color. Only the window indicated to be the next target could be destroyed with the ball. The order in which the windows were to be hit was defined randomly and the same for everyone. A panel displayed the attempts and time needed so far. The player behavior and performance were logged.

3.3 Procedure

The experiment was conducted in the spacious clubhouse of a local soccer club in compliance with the infection control regulations at the time. In the beginning, the participants were briefed and signed a consent form. After a Vive Tracker was attached to their right foot with elastic bands, the position of the virtual foot was calibrated in Unity. Then, the players were given a head-mounted display and a Vive controller. In the VE, the players saw the wall with the targets and 11.22 m away from it the ball in front of them. Approaching the ball, they had the chance to practice with up to 30 shots but could end the practice phase early. Then, the shooting task started in which eight adjacent windows had to be hit in the given order. The players had as many attempts and as much time as they needed for completing the task. The balls always spawned in the same location. After completion, the participants left VR and filled in the following questionnaires on a tablet:

a) Demographic data and prior experience in soccer, VR, and gaming,

b) The applicable subscales *Involvement/Control*, *Naturalness*, and *Interface Quality* from the Presence Questionnaire (PQ) by Witmer and Singer [39],

c) The NASA raw Task Load Index (TLX) [13],

d) Seven custom items on the subjective control over the ball as detailed in Appendix A.1, and

e) Five custom items on body ownership as detailed in Appendix A.2.

We could not use validated questionnaires on body ownership for this study, because some items cannot be applied to the condition with invisible foot. Our custom questions are inspired from standardized body ownership scales [9, 18] and adapted to our scenario. Concluding the experiment, we conducted a semi-structured interview with the players. We asked about their feel for the ball, different strategies of kicking, the interaction fidelity, their satisfaction with the achieved performance, previous soccer experience, and if they would like to play again. The interview guideline is detailed in Appendix A.3. The sessions took about 45 minutes with about 15 minutes in VR. The players were compensated with snacks and beverages for their participation.

3.4 Data Analysis

All statistical tests were calculated with an alpha level of .05 and assume an advantage of displaying the foot as alternative hypotheses. For the metric data on times and attempts, we applied Student *t* tests for independent samples. It was approximately normally distributed as assessed by Shapiro Wilk tests ($p = [0.096 .. 0.694]$). The homogeneity of variances was confirmed with Levene's tests for all metric data ($p = [0.115 .. 0.719]$) except for the number of attempts ($F = 11.054$, $p = .003$) for which Welch's *t* test was used, consequently. The distribution of the ordinal data from the questionnaires was compared with non-parametric Mann-Whitney *U* tests.

The custom questions on the subjective ball control and the body ownership have not been validated as standardized scales. Therefore, we conducted a reliability analysis to check for the internal consistency of the items. The tests yielded Cronbach's Alpha values of $\alpha_{BC} = .88$ for the ball control items and $\alpha_{BO} = .89$ for the body ownership items, strongly indicating that the items measure the same underlying concept, respectively. We assessed potential influences of sample characteristics as experience or demographics with Pearson's correlations if both variables are metric and Spearman's correlations, otherwise, for ordinal data. We used *t* and *U* tests for the nominal factor gender. The qualitative data was analyzed with an inductive thematic analysis.

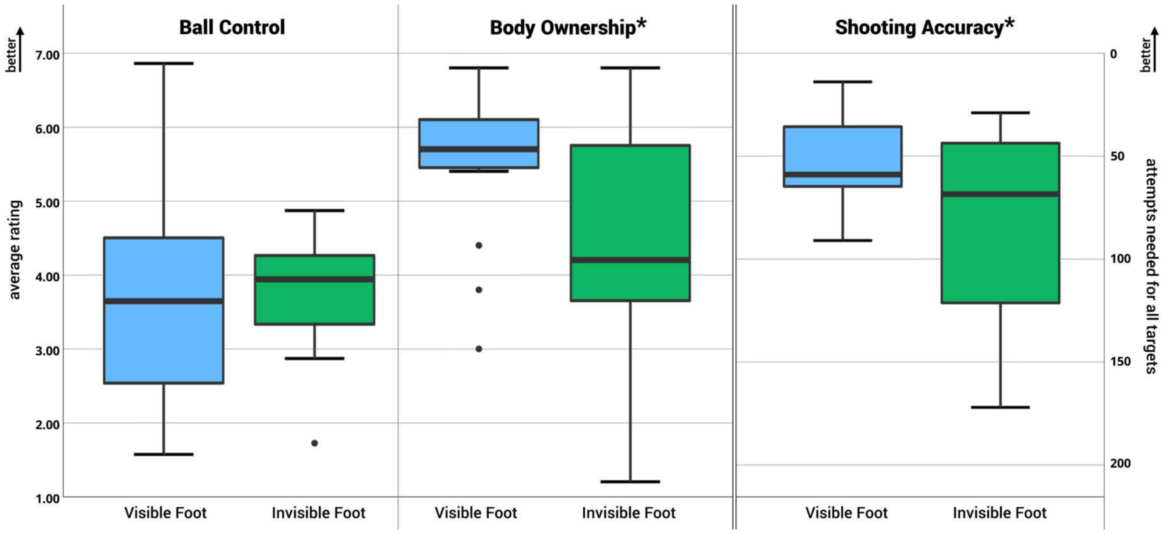


Figure 3: Box plots comparing the distribution of ratings on ball control and body ownership as well as the shooting accuracy between the two conditions. Please note the two different scales. Significant group mean differences with $p < .05$ are marked with an asterisk.

4 RESULTS

We compared the logged performance data between the two conditions with visible foot (+F) and invisible foot (–F) as well as the user ratings from the questionnaires. Further, we report what we learned about the participants' experiences in the interviews.

4.1 Shooting Performance

In the experiment, we measured how fast and accurately the players were able to hit eight targets. It took them a similar amount of **time** in both groups to complete the task. On average, players without virtual representation of their foot needed 77 s longer (7:38 min), but this difference is not significant ($p = .105$), as can be seen in Table 1 – along with all further descriptive and test statistics.

Regarding the **accuracy** of the shots, players with visible foot performed better in the task. They needed an average of only 6.56 attempts per target, compared to 10.33 attempts in the –F condition. For all eight targets, we observe a significant difference of 30.14 attempts which is 57.5% more on average and corresponds to a large effect size [8] with $d_{Cohen} = 0.82$, $p = .022$. While it was possible for some players with invisible foot to achieve a similar performance as with visible foot, there were substantially more participants in the –F condition with less accuracy – with up to 172 attempts, as illustrated in Figure 3. The best player hit all 8 targets with only 14 shots and was in the condition with foot visualization. As Levene's test confirms, the variance in the –F condition was significantly larger ($F(1) = 11.054, p = .003$). The prior experience in playing soccer had no measurable influence on the accuracy ($\rho = -.169$, $p = .390$) or time needed ($\rho = -.229$, $p = .241$), and neither did VR experience, gaming experience, age, or gender (all with $p \geq .296$).

From the accuracy and time, we can assess the **shot rate**, i.e., how often the participants kicked the ball per minute. On average, the players who could see their virtual foot shot with a lower frequency (8.45 shots per minute) than the group with invisible foot (10.48). Hence, they took an average of 1.85 s longer to take a shot, unlike what was hypothesized.

4.2 User Ratings

After the task, we asked our participants to fill in questionnaires about their experience and rate it on a Likert scale from 1 to 7, except for the TLX rated from 10 to 100. In the **Presence** Questionnaire [39], we did not find group differences in the ratings of any of the tested subscales *Involvement/Control*, *Naturalness*, and *Interface Quality*. Similarly, the NASA raw **Task Load Index** [13] showed no significant difference between the groups with average scores of 31.

The assessments of **body ownership** reveal a significant difference between the groups with higher ratings by players who could see their foot ($M_{+F} = 5.51$) compared to the condition with invisible foot ($M_{-F} = 4.44$). This is a medium to large effect with $d_{Cohen} = 0.71$, $p = .040$. There is a large spread of opinions among the players with no foot representation on how high the body ownership was, compared to more consistent ratings in the +F condition. A heterogeneity of variances is suggested by Levene's test ($F(1) = 3.047, p = .093$).

The ratings for aspects of **ball control** are on average equal in both conditions with moderate scores ($M_{+F} = 3.72$, $M_{-F} = 3.74$, $p = .600$). Particularly low are the ratings for how precisely the players could control the force of a shot with an average rating of 2.68. Conversely to the body ownership assessments, we observe a large scattering of ball control ratings by players with visible foot indicating diverse perceptions in the +F condition, compared to more uniform assessments by players with invisible foot. Again, Levene's test raises doubts about the homogeneity of variances between the conditions ($F(1) = 3.923, p = .058$). The distributions of the ball control and body ownership scores are visualized in Figure 3.

Neither the soccer, VR, and gaming experiences nor the age or gender of the participants had any significant influence on the questionnaire results (all with $p \geq .125$), with one exception: the more experience the players had in soccer, the more the interface was reported to interfere with their performance ($\rho = -.392$, $p = .039$).

4.3 Feedback from the Players

In the interviews after the experiment, we noticed great ambitions by the players to improve their kicking skills in our simulation. In particular, experienced soccer players wanted to practice more. Participants from both conditions asked if they may continue playing. We observed that participants who achieved a good feel for the ball were satisfied with the system but tended to criticize their own skills. In contrast, players who did not manage to acquire intuitive control over the ball looked for flaws in the system. Some participants complained that the ball sometimes only rolled ponderously or in an unintended direction, although they had the impression that they hit it correctly. The players described various strategies for improving their shooting abilities similar to the learning process of soccer beginners getting familiar with kicking a real ball. For instance, they tried hitting the virtual ball in different locations, hitting it from different angles, or using the heel of the foot for kicking.

When being asked about the visualization of the foot, most participants who could not see a virtual foot during the task explained that they did not miss it. Some suspected that it might have even been distracting. In the condition with visible foot, some players assumed that they would not have needed the visualization and would have performed well without it. In one case, the player only realized at the end of the task that he could in fact see his virtual foot. Further, only one participant suggested displaying the left foot as well to get a better idea of how he stands.

5 Discussion

From our results, we learn that visualizing the virtual foot brings benefits for shooting targets with a virtual ball. Although showing a virtual soccer shoe did not significantly improve how fast the players hit the targets, it has increased the accuracy of their shots. Players without visual reference aimed significantly worse and needed inordinately many attempts: up to 21.5 shots per target. Showing the virtual foot enabled the players to hit the targets more reliably. The best performing player could see the foot and needed only 1.75 shots per target. Hiding the virtual foot resulted in 57.5% additional attempts on average and a larger variance. These findings support our hypothesis H1 on advantages for the kicking performance. The increased accuracy in foot actions is in line with findings by Kosmalla et al. [17] on visualized feet when climbing in VR who demonstrated a reduction of self-reported stepping errors. Previous studies found similar effects for typing in VR where hiding hand visualizations resulted in high error rates [10] and in particular for inexperienced typists required more time and error corrections [16].

On the other hand, the average perceived control over the ball, which we interpret as an indicator of agency over the feet, yielded no difference between the group means, unlike what was assumed with H4. Interestingly, the visible foot caused a large variance among the participants regarding how high their subjective ball control was. We suspect that for some players, calibration inaccuracies might have caused unexpected ball behavior that did not match the observed action and thus lowered the perceived control, while for others with good calibration the visibility in fact enhanced the ball control. In contrast, participants who could not see their foot had to rely on proprioceptive cues and intuition independent of calibration offsets leading to more uniform impressions of ball control.

Professional soccer players do not look at the ball or their foot while taking a shot. However, kicking a ball in reality renders a detailed haptic response on the foot. For developing a feel for the ball, this lack of haptic sensation can be a challenging gap in the feedback loop while learning how best to hit the ball, which can only be compensated with proprioceptive or visual cues. Most of the players with invisible foot explained in the interviews that they did not miss a virtual representation of their foot. They were largely in agreement on their level of ball control but performed worse. In contrast, players who were able to see their foot could not agree less on their subjective ball control, but in fact shot more accurately. The impact of displaying a shoe model might take effect on a subconscious level. Indeed, one player only became aware of the foot visualization at the end of the session. Still, the visual feedback might have been effective in improving this player's estimate how to best hit the ball. By comparing the assumed position of the foot with the visual information on its actual position, the player could readjust the kicking movement without realizing it. This subconscious feedback loop and motor recalibration seem especially plausible with a high body ownership.

Indeed, the visibility of the foot had a significant impact on the perceived body ownership of the players. Being able to see the foot they shoot with enhanced their body ownership on average, which confirms hypothesis H5. Additionally, we observed a similar scattering phenomenon as with the ball control. While players with visible foot reported consistently high scores, the variation in the answers from the invisible foot condition shows large disagreement within the group, with a wide range of body ownership ratings. Some players seemed to have difficulties relating to a virtual foot that they cannot see despite being able to impact the virtual environment with it. For others, the effectiveness of their actions was sufficient for owning the virtual foot.

Other than expected, our investigation showed no advantage from the foot visibility in terms of task load (H2) or presence (H3). Previous work has shown a higher workload and lower presence when hiding the hands for a typing task [16]. Also in tasks that involve the feet, showing a self-avatar was found to be an important contributor to the plausibility illusion [29] and to increase presence [22]. This discrepancy could be attributed to the visual focus of the task that was rather on the targets than on the foot visualization. We would expect to see effects on task load and presence for activities that focus visual attention on the feet.

To our surprise, we also observed no influence from prior soccer experience on the performance, ball control, or body ownership in our simulation. From amateur to professional, everybody needed to get familiar with the virtual ball and adapt to its behavior. Only regarding the interface quality, experienced players diverged in their assessment and gave worse ratings for how much the system interfered with their kicking performance, independent from the condition. As a possible explanation, we suspect that the professionals are used to a better shooting performance and therefore felt particularly limited in their capabilities in the simulation. The performance of throwing a ball in reality is twice as accurate as in VR according to a study by Zindulka et al. [41]. There could be a similar limitation in kicking as in throwing in VR. After the experiment, many players asked to continue playing and practicing with our prototype. They expressed having a lot of fun in shooting the targets and wanted to further improve their virtual kicking skills.

5.1 Limitations

Some participants acquired a precise and intuitive feel for the ball yielding in performances with only few attempts, whereas other participants struggled with the control over the ball and criticized its unrealistic behavior. In particular, how precisely they could control the force of the shot was rated poorly by the players. Therefore, the collision detection or force transfer might be limiting factors of the tested prototype. This applies especially for powerful shots according to observations and feedback. One explanation could be that if the foot motion were too fast, the collision occurring between frames might not have been detected correctly and thus a wrong speed profile was calculated. While this issue affected both conditions equally it might account for some of the noise in the data and conceal effects that would be traceable otherwise.

Further, we assume that the calibration was not ideal for some participants as they developed coping strategies compensating a potential offset. Since the foot model and collider dimensions were not scaled to match the shoe size of the players, the offset varied depending on the size of their feet. Although 30 practice shots allowed for adapting to calibration shortcomings which affected both conditions equally, we cannot exclude confounding effects due to noisy data. As suggested in the literature [1, 23, 27], there might also be an influence of the break in presence from removing the VR headset on the subjective measures in both conditions, which could be an additional cause of high variances in the data.

We can further assume that the sample size of 28 participants in combination with the between-group design was a limiting aspect in this study. It is conceivable that with a larger sample, further differences between the groups would have become apparent, especially regarding the required time, although further recruitment would have made it difficult to ensure a sample covering diverse soccer experiences. Lastly, it might have been beneficial for the players' body ownership if both feet were visible and if more degrees of freedom were mapped by tracking the ankle joint. Even though this requires additional time in putting on and calibrating at least four trackers, it should be worth the effort in future studies.

5.2 Future Work

Based on our data and in line with comments from the interviews, we suspect subconscious processes being involved that require more purposeful investigations of target shooting. For instance, we suggest capturing the direction in which the player looks and whether the foot is within the field of view while taking a shot as well as the distance by how much a target was missed. A comparison of real and virtual ball shooting, as has been done for throwing [41], would help to understand the unique characteristics of kicking and foot-eye coordination in VR. Moreover, considering related research on hand visualizations, we expected an increase of presence and agency over the ball when displaying the foot. One explanation for the deviating results of our study could be that the player's visual focus in penalty shooting is primarily on the goal, less on the kicking action itself. It should be interesting to compare the results presented here to future research on foot interaction with the action in the center of visual attention and more fine motor demanding ball handling, such as balancing or juggling the ball with the feet.

Also, other aspects of playing soccer could be investigated to gain a deeper understanding of foot–ball interaction in VR and the influence of limb visualization and embodiment, such as passing the soccer ball to other players in a multi-user scenario or dribbling it while moving along. Equivalent to similar work on hand interaction, it would also be interesting to compare various visual representations of the foot, such as abstract and realistic foot avatars. Finally, enriching the kicking action with haptic feedback, such as force feedback to the foot at collision with the ball, might benefit the player's presence and body ownership. A nuanced haptic actuator indicating the location of impact along the inside of the foot could have a strong impact on motor learning and improve the performance. Combining foot with hand visualizations and haptic feedback for full-body activities, such as in martial arts or yoga, will allow integrating the role of visual foot feedback in embodiment models and skill acquisition for VR sports.

6 Conclusion

The presented study provides insights into the role of foot visibility when interacting with a virtual ball for a central component of playing soccer in VR: the targeted kicking of a ball in a penalty shooting task from a stationary position. We found that seeing one's foot while shooting a ball can significantly decreases the required number of attempts for hitting the targets. Although this suggests an advantage for controlling the aiming the shots, the subjective control over the virtual ball did not benefit from the foot visibility overall. Instead, it caused a high variance within the group's self-reports. We suspect subconscious processes with high sensitivity towards calibration inaccuracies causing visual mismatch.

While displaying the foot did not decrease the task load or improve players' presence, our findings show significantly higher and more unanimous body ownership assessments with a visible foot. Further studies are required to understand the effects of foot visualization on body ownership and agency that came into play in this first evaluation. On the journey towards expanding full-body tracking and embodied interaction in VR, more research on foot visualization will enable high-fidelity interactions with the feet in sports and beyond.

Acknowledgments

This work was partially funded by the Klaus Tschira Foundation, the KI-SIGS Project "AP 390" of the German Federal Ministry of Education and Research (BMBF), and the German Research Foundation (DFG) as part of the Collaborative Research Center (SFB) 1320 *EASE – Everyday Activity Science and Engineering*, University of Bremen (`www.easecrc.org`).

References

[1] D. Alexandrovsky, S. Putze, M. Bonfert, S. Höffner, P. Michelmann, D. Wenig, R. Malaka, and J. D. Smeddinck. *Examining Design Choices of Questionnaires in VR User Studies*, p. 1–21. Association for Computing Machinery, New York, NY, USA, 2020.

[2] F. Anderson, M. Annett, and W. Bischof. Lean on wii: physical rehabilitation with virtual reality wii peripherals. *Studies in health technology and informatics*, 154:229–34, 2010.

[3] F. Argelaguet, L. Hoyet, M. Trico, and A. Lecuyer. The role of interaction in virtual embodiment: Effects of the virtual hand representation. In *2016 IEEE Virtual Reality (VR)*, pp. 3–10, 2016. doi: 10.1109/VR.2016.7504682

[4] L. Bozgeyikli and E. Bozgeyikli. Tangiball: Dynamic embodied tangible interaction with a ball in virtual reality. In *Companion Publication of the 2019 on Designing Interactive Systems Conference 2019 Companion*, DIS '19 Companion, p. 135–140. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3301019.3323904

[5] R. Canales, A. Normoyle, Y. Sun, Y. Ye, M. D. Luca, and S. Jörg. Virtual grasping feedback and virtual hand ownership. In *ACM Symposium on Applied Perception 2019*, SAP '19. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3343036.3343132

[6] M. Carrozzino, G. Avveduto, F. Tecchia, P. Gurevich, and B. Cohen. Navigating immersive virtual environments through a foot controller. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology*, VRST '14, p. 23–26. Association for Computing Machinery, New York, NY, USA, 2014. doi: 10.1145/2671015.2671121

[7] S. Cmentowski, A. Krekhov, A. Zenner, D. Kucharski, and J. Krüger. Towards sneaking as a playful input modality for virtual environments. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 473–482, 2021. doi: 10.1109/VR50410.2021.00071

[8] J. Cohen. *Statistical Power Analysis for the Behavioral Sciences*. Erlbaum, Hillsdale, NJ, 2nd ed., 1988.

[9] M. Gonzalez-Franco and T. C. Peck. Avatar embodiment. towards a standardized questionnaire. *Frontiers in Robotics and AI*, 5:74, 2018. doi: 10.3389/frobt.2018.00074

[10] J. Grubert, L. Witzani, E. Ofek, M. Pahud, M. Kranz, and P. O. Kristensson. Effects of hand representations for typing in virtual reality. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 151–158, 2018. doi: 10.1109/VR.2018.8446250

[11] T. Han, J. Alexander, A. Karnik, P. Irani, and S. Subramanian. Kick: investigating the use of kick gestures for mobile interactions. In *MobileHCI '11 Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, pp. 29–32. Association for Computing Machinery, USA United States, Aug. 2011. Mobile HCI 2011 ; Conference date: 30-08-2011 Through 02-09-2011. doi: 10.1145/2037373.2037379

[12] D. J. Harris, G. Buckingham, M. R. Wilson, J. Brookes, F. Mushtaq, M. Mon-Williams, and S. J. Vine. Exploring sensorimotor performance and user experience within a virtual reality golf putting simulator. *Virtual Reality*, pp. 1–8, 2020.

[13] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati, eds., *Human Mental Workload*, vol. 52 of *Advances in Psychology*, pp. 139–183. North-Holland, 1988. doi: 10.1016/S0166-4115(08)62386-9

[14] M. Hoppe, J. Karolus, F. Dietz, P. W. Woundefinedniak, A. Schmidt, and T.-K. Machulla. *VRsneaky: Increasing Presence in VR Through Gait-Aware Auditory Feedback*, p. 1–9. Association for Computing Machinery, New York, NY, USA, 2019.

[15] K. Kilteni, R. Groten, and M. Slater. The sense of embodiment in virtual reality. *Presence: Teleoper. Virtual Environ.*, 21(4):373–387, Dec. 2012. doi: 10.1162/PRES_a_00124

[16] P. Knierim, V. Schwind, A. M. Feit, F. Nieuwenhuizen, and N. Henze. *Physical Keyboards in Virtual Reality: Analysis of Typing Performance and Effects of Avatar Hands*, p. 1–9. Association for Computing Machinery, New York, NY, USA, 2018.

[17] F. Kosmalla, A. Zenner, C. Tasch, F. Daiber, and A. Krüger. The importance of virtual hands and feet for virtual reality climbing. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20, p. 1–8. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3334480.3383067

[18] M. R. Longo, F. Schüür, M. P. Kammers, M. Tsakiris, and P. Haggard. What is embodiment? a psychometric approach. *Cognition*, 107(3):978–998, 2008. doi: 10.1016/j.cognition.2007.12.004

[19] K. Ma and B. Hommel. The role of agency for perceived ownership in the virtual hand illusion. *Consciousness and Cognition*, 36:277–288, 2015. doi: 10.1016/j.concog.2015.07.008

[20] J. Montes and P. Figueroa. Vr salsa: Learning to dance in virtual reality. In *Proceedings of the IX Latin American Conference on Human Computer Interaction*, CLIHC '19. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3358961.3358969

[21] V. Paelke, C. Reimann, and D. Stichling. Foot-based mobile interaction with games. In *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*, ACE '04, p. 321–324. Association for Computing Machinery, New York, NY, USA, 2004. doi: 10.1145/1067343.1067390

[22] Y. Pan and A. Steed. How foot tracking matters: The impact of an animated self-avatar on interaction, embodiment and presence in shared virtual environments. *Frontiers in Robotics and AI*, 6:104, 2019.

[23] S. Putze, D. Alexandrovsky, F. Putze, S. Höffner, J. D. Smeddinck, and R. Malaka. *Breaking The Experience: Effects of Questionnaires in VR User Studies*, p. 1–15. Association for Computing Machinery, New York, NY, USA, 2020.

[24] D. Roth and M. E. Latoschik. Construction of the virtual embodiment questionnaire (veq). *IEEE Transactions on Visualization and Computer Graphics*, 26(12):3546–3556, 2020. doi: 10.1109/TVCG.2020.3023603

[25] A. Rovers and H. Essen, van. Guidelines for haptic interpersonal communication applications : an exploration of foot interaction styles. *Virtual Reality*, 9(2-3):177–191, 2006. doi: 10.1007/s10055-005-0016-0

[26] 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*, CHI '19, p. 1–13. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3290605.3300413

[27] V. Schwind, P. Knierim, N. Haas, and N. Henze. *Using Presence Questionnaires in Virtual Reality*, p. 1–12. Association for Computing Machinery, New York, NY, USA, 2019.

[28] V. Schwind, P. Knierim, C. Tasci, P. Franczak, N. Haas, and N. Henze. *"These Are Not My Hands!": Effect of Gender on the Perception of Avatar Hands in Virtual Reality*, p. 1577–1582. Association for Computing Machinery, New York, NY, USA, 2017.

[29] R. Skarbez, S. Neyret, F. P. Brooks, M. Slater, and M. C. Whitton. A psychophysical experiment regarding components of the plausibility illusion. *IEEE Transactions on Visualization and Computer Graphics*, 23(4):1369–1378, 2017. doi: 10.1109/TVCG.2017.2657158

[30] M. Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364:3549 – 3557, 2009.

[31] H. Son, H. Gil, S. Byeon, S.-Y. Kim, and J. R. Kim. Realwalk: Feeling ground surfaces while walking in virtual reality. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI EA '18, p. 1–4. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3170427.3186474

[32] P. Strohmeier, S. Güngör, L. Herres, D. Gudea, B. Fruchard, and J. Steimle. *BARefoot: Generating Virtual Materials Using Motion Coupled Vibration in Shoes*, p. 579–593. Association for Computing Machinery, New York, NY, USA, 2020.

[33] A. Tajadura-Jiménez, M. Basia, O. Deroy, M. Fairhurst, N. Marquardt, and N. Bianchi-Berthouze. *As Light as Your Footsteps: Altering Walking Sounds to Change Perceived Body Weight, Emotional State and Gait*, p. 2943–2952. Association for Computing Machinery, New York, NY, USA, 2015.

[34] L. Terziman, M. Marchal, F. Multon, B. Arnaldi, and A. Lecuyer. The king-kong effects: Improving sensation of walking in vr with visual and tactile vibrations at each step. In *2012 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 19–26, 2012. doi: 10.1109/3DUI.2012.6184179

[35] M. Tiator, B. Fischer, L. Gerhardt, D. Nowottnik, H. Preu, and C. Geiger. Cliffhanger-vr. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 1–2, 2018. doi: 10.1109/VR.2018.8446506

[36] D. Valkov, F. Steinicke, G. Bruder, and K. H. Hinrichs. Traveling in 3d virtual environments with foot gestures and a multi-touch enabled wim. In *Proceedings of the Virtual Reality International Conference (VRIC)*, pp. 171–180, 2010.

[37] E. Velloso, D. Schmidt, J. Alexander, H. Gellersen, and A. Bulling. The feet in human–computer interaction: A survey of foot-based interaction. *ACM Comput. Surv.*, 48(2), Sept. 2015. doi: 10.1145/2816455

[38] J. von Willich, M. Schmitz, F. Müller, D. Schmitt, and M. Mühlhäuser. Podoportation: Foot-based locomotion in virtual reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, p. 1–14. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3313831.3376626

[39] B. G. Witmer and M. J. Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoper. Virtual Environ.*, 7(3):225–240, June 1998. doi: 10.1162/105474698565686

[40] P. W. Woźniak, P. Knierim, and M. Hoppe. Health, fun, and engagement: Computing technologies that support physical activity. *XRDS*, 25(4):16–20, July 2019. doi: 10.1145/3331014

[41] T. Zindulka, M. Bachynskyi, and J. Müller. Performance and experience of throwing in virtual reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, p. 1–8. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3313831.3376639

A APPENDIX

Twelve custom questionnaire items on the subjective feel for the ball and on body ownership were rated on a Likert scale from 1 (not at all) to 7 (completely). The last item on ball control was erroneously placed in the questionnaire on body ownership and is therefore phrased as a statement. The reported calculations account for the correct attribution as listed in the following.

A.1 Items on Ball Control

1. How precisely were you able to control the direction of the shot?
2. How precisely were you able to control the force of the shot?
3. How realistic was the shot?
4. Were you able to aim the ball?
5. Did the ball hit the spot you aimed for?
6. Did the ball behave the way you wanted it to?
7. I felt the force of my biological foot transferred to the ball.

A.2 Items on Body Ownership

1. I was aware of the location of my biological foot.
2. I was aware of the location of my virtual foot.
3. I felt that my biological foot was present in the virtual environment.
4. I knew exactly where my foot was in the virtual environment.
5. I was confused about my foot in the virtual environment. *[reversed]*

A.3 Interview Guideline

The following questions and possible follow-up questions served as the basis for the semi-structured interview and were flexibly adapted depending on what the participants shared.

- Would you like to play again?
- Were you able to control the ball well?
  - Did you develop a strategy for targeting your shots?
  - Did you try different strategies? What worked well and what did not?
- How realistic did it feel to shoot the ball?
  - How would you describe your feel for the ball?
  - How would you describe the behavior of the ball?
- Were you satisfied with the result you achieved?
- Do you think, previous experience in soccer is an advantage in this game?
- Would you like to share anything else?
  - Do you have any suggestions for improvement?
  - Do you have other impressions that have not yet been addressed?

Publication P7

VRisbee: How Hand Visibility Impacts Throwing Accuracy and Experience in Virtual Reality

Malte Borgwardt, Jonas Boueke, María Fernanda Sanabria, Michael Bonfert, and Robert Porzel

This publication investigates how the visibility of virtual hands can support players when throwing disks. We developed a Frisbee simulation in VR and asked 29 study participants to hit a target in a between-group experiment. The results show a subtle advantage of hand visibility in terms of accuracy. Visible hands further improved the impression of realism and subjective disc control. The evidence of a change in body ownership was inconclusive.

My contribution: I contributed to the formulation of the research goals (*conceptualization*) and the study design (*methodology*). I made an equal contribution to the statistical analysis (*formal analysis*) as JB. I made a major contribution to writing and editing the manuscript (*writing – review & editing*). Together with RP, I supervised the project (*supervision*).

Published in the *Extended Abstracts of the 2023 ACM Conference on Human Factors in Computing Systems (CHI EA '23)*

VRisbee: How Hand Visibility Impacts Throwing Accuracy and Experience in Virtual Reality

Malte Borgwardt
borgward@uni-bremen.de
University of Bremen
Bremen, Bremen, Germany

Jonas Boueke
jboueke@uni-bremen.de
University of Bremen
Bremen, Bremen, Germany

María Fernanda Sanabria
mariafer@uni-bremen.de
University of Bremen
Bremen, Bremen, Germany

Michael Bonfert
University of Bremen, Digital Media Lab
Bremen, Bremen, Germany
bonfert@uni-bremen.de

Robert Porzel
University of Bremen, Digital Media Lab
Bremen, Bremen, Germany
porzel@uni-bremen.de

Figure 1: Scene of the virtual reality Frisbee simulation developed for the study

ABSTRACT

Hand interaction plays a key role in virtual reality (VR) sports. While in reality, athletes mostly rely on haptic perception when holding and throwing objects, these sensational cues can be missing or differ in virtual environments. In this work, we investigated how the visibility of a virtual hand can support players when throwing and what impact it has on the overall experience. We developed a Frisbee simulation in VR and asked 29 study participants to hit a target. We measured the throwing accuracy and self-reports of presence, disc control, and body ownership. The results show a subtle advantage of hand visibility in terms of accuracy. Visible hands further improved the subjective impression of realism, body ownership and subjective control over the disc.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; **Virtual reality**.

CHI EA '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9422-2/23/04.
https://doi.org/10.1145/3544549.3585868

KEYWORDS

virtual reality, throwing, embodiment, self-avatar, sports

ACM Reference Format:
Malte Borgwardt, Jonas Boueke, María Fernanda Sanabria, Michael Bonfert, and Robert Porzel. 2023. VRisbee: How Hand Visibility Impacts Throwing Accuracy and Experience in Virtual Reality. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3544549.3585868

1 INTRODUCTION

Every sport requires specific technical abilities to achieve the main objective, for example, to throw a Frisbee accurately is required for *Ultimate* or *Disc Golf*. Frisbee throwing skills and tactics differ from other sports, as throwing a disc relies on other physical principles than throwing a ball or a dart, for example. A disc travels on a flat trajectory stabilized by spin and lift forces that are highly dependent on the disc's orientation. Since virtual reality (VR) technology evolved and gained popularity, it can help improve or practise skills that can be more difficult to replicate in a real-world setting [21]. Approaches range from learning musical instruments, training astronauts, and educating in the military or medical fields, to applications in sports. For skill acquisition in virtual simulations, the learned skill needs to be transferable to reality, e.g., for throwing real darts [16, 21]. However, especially for throwing, the transfer between the virtual and the real world comes with challenges and currently requires trade-offs, e.g., in terms of distance perception [4] or aiming [23].

Throwing a disc additionally requires players to consider the weight and shape of the disc as well as external factors such as wind through haptic stimuli to anticipate the possible outcome of the throw. In current consumer VR systems with controllers or hand tracking, these cues are not available to the player. How can we still enable players to accurately throw a disc and control its flight?

We investigated supporting Frisbee throwing with a visual hand representation. In VR setups, users are often supported with visual self-representations such as full or partial self-avatars. Studies found significant benefits of hand visibility for tasks like typing on a virtual keyboard [8] or picking and placing objects [1, 5] as it has a positive influence on players' presence and body ownership in the virtual world. This study aims to understand if hand visibility has an impact on the performance and sense of embodiment when throwing a disc in VR. Therefore, we investigate the following hypotheses: **H1** The visibility of the virtual hands improves the accuracy of throwing a virtual Frisbee; **H2** Hand visibility increases perceived presence; **H3** Hand visibility increases the sense of body ownership; **H4** Hand visibility increases the subjective control over the virtual Frisbee. We report on the design and results of a user study ($N = 29$) and discuss the outcome. Further, we contribute a prototype with realistic Frisbee trajectory physics.

2 RELATED WORK

Different research areas are related to this interdisciplinary topic, such as human-computer interaction, sports, computer science, and psychology. We present related work concerning aspects of embodiment and its effects on interactions and sports in VR.

2.1 Aspects of Embodiment

Kilteni et al. [12] explain the sense of embodiment as a combination of three concepts: self-location, sense of agency, and sense of body ownership. For a plausible sense of self-location, it is important that the biological and virtual bodies are perceived to be in the same space [5]. A high sense of agency requires having motor control over the virtual body's actions [12]. The sense of body ownership refers to "one's self-attribution of a body" [12] and implies that the body is the agent of the experienced sensations [7]. The user's perceived level of embodiment can strongly affect the VR experience. The concept is linked to the user's presence in a virtual environment [18, 20] and might therefore determine how well experiences and skill acquisition from the virtual world can be transferred to the real world.

There is a growing body of literature on how VR training in sports can benefit performance in reality. In a study by Pastel et al. [19], participants completed three tasks in both virtual and real environments: balance, grasping and throwing. Its results showed that full body visualisation led to better performances, especially if compared with no body visualisation at all, and that a realistic virtual body helps to "limit differences between the real world and virtual reality and to ensure quite natural body perception" [19]. Research on body ownership concerning the user's virtual hands by Canales et al. [5] investigated the influence of different mappings on a manipulation task in virtual reality. They found that participants preferred adjusted visualisations that prevented the hand from intersecting with objects, even though displaying the original position improved performance, which also induced higher ownership. Similarly, Ma and Hommel [15] showed the role of agency for perceived hand ownership. The degree to which the virtual agent could be controlled by people's own movements was at least as important as a realistic appearance of the hand when it comes to body ownership. Further, a study by Grubert et al. [8] evaluated different hand and fingertip representations to find the best visual support for extensive text entry. Minimalist visualisation of only the fingertips proved to be the preferred option, with a high input rate and low error rate. This outperformed the high-fidelity variant of a full 3D representation of the hand, which in turn induced higher body presence at the cost of obscuring the keyboard. Therefore, it might not be possible to generalize the ideal hand representation for any use case, and rather depend on the scenario, the goal, the input and output devices, and individual preferences. This is also supported by the findings of Kocur et al. [13], who conducted a study to investigate the impact of missing individual fingers in virtual interactive scenarios. They found, that a missing virtual representation of the index finger, which was by far the most used by the participants, led to the strongest drop in body presence. Comparing this to the effect of replacing the virtual hand by an abstract one, it showed, that the usefulness of a hand representation has a much higher influence than realism.

2.2 Sports in Virtual Environments

With innovations in sensor technology enabling position, motion and biofeedback detection for complex posture recognition or physiological measurements, doing sports in virtual environments opens up new opportunities. Some systems are particularly designed for coaching and training people, both for supporting physical activity and providing motivation [2].

Neumann et al. [17] concluded in their systematic review of interactive sports applications in VR that while "the majority of research has been conducted on endurance sports, such as running, cycling, and rowing, more research is required to examine the use of interactive VR in skill-based sports" [17]. This is particularly challenging as precise movements and actions, such as throwing, are difficult to capture and transfer accurately to virtual actions. For example, Harris et al. [9] investigated the performance of golf putting in VR and compared it to the performances in the real world. Their study showed that expert players also outperformed novices in the virtual world, suggesting that a VR simulation can achieve a high level of fidelity so that it requires skills comparable to the real world. As another example, a study by Tirp et al. [21] on skill transfer of throwing darts compared the accuracy and quiet eye duration between real and virtual throwing. The results showed that participants with virtual training threw with the highest accuracy which implies a high skill transferability. However, these results conflict with findings by Drew et al. [6]. In a similar study, the participants who practised dart throwing in VR performed worse than the real-life training group. Furthermore, a discrepancy between real and virtual throwing is demonstrated in an investigation by Zindulka et al. [23] who compared overhand and underhand ball throws. Their results showed a significantly worse performance in VR compared to real throwing regarding accuracy and precision, with pronounced scattering especially in the vertical axis. These findings could be explained by using controllers for throwing. Timing the release of the object with the controller's trigger button seems challenging compared to releasing a physical ball from the fingers. Only small deviations from the optimal release moment can lead to a strong impact on the trajectory of the virtual object. Although a disc with a horizontal trajectory is thrown in our study, the limited control over the release point might also be a limiting factor of our controller-based system.

Concerning the embodiment during sports in VR, Kosmalla et al. [14] tested the influence of limb visualisation on the climbing experience in VR. The climbers' hands end feet were tracked and the virtual route matched the physical wall. The authors reported that the perceived stepping errors were significantly higher without foot visualisation. However, the perceived gripping errors were not affected by showing or hiding the virtual hands. With more experience in climbing, the participants were able to compensate for the lack of hand visibility by only relying on proprioceptive cues. Similarly, Bonfert et al. [3] analysed the influence of foot visibility on a kicking task in VR. They found that the visible foot improved the accuracy of the shot, even though the subjective control over the virtual ball did not improve. The self-reported body ownership was reported higher with the visible foot. While these studies show that the visibility of the hands and especially feet can improve the performance and user experience in VR sports, they do not address the influence of hand visibility in throwing tasks. We are trying to fill this gap with our study using the system we specifically developed for this purpose.

3 SYSTEM DESIGN

To learn about the influence of hand visibility in VR sports, we created a VR game that simulates the throwing of a Frisbee. It was built for the Meta Quest 1 with its standard Touch Controllers. The object interaction was based on the physics engine and the XR Interaction Toolkit of Unity 2021.3.3f1. The application allowed grasping discs and launching them with the speed and direction of the controller motion. However, as the disc trajectory is not parabolic, the hovering and curve are not calculated realistically by the physics engine and had to be addressed manually. The implementation of the Frisbee flying physics is grounded in the works by Hubbard and Hummel [10] and Hummel [11] who developed a numerical simulation and estimated the model's aerodynamic coefficients to match the trajectory of a real disc. We adjusted the virtual disc's parameters of translational velocity, angular velocity, and tilt in Unity according to the estimated forces of the initial impulse, lift, drag, and torques. One part of the calculation addresses the disc's lift (L) and drag (D), they are calculated as shown in Equation 1. Both are dependent on the plan-form area of the disc (A), the velocity (v), the air density (ρ) as well as a lift coefficient (C_l) for lift and a drag coefficient (C_d) for drag. C_l is linearly dependent on α while C_d is dependent on α^2 [10].

$$L = \frac{1}{2} C_l A \rho v^2, \quad D = \frac{1}{2} C_d A \rho v^2 \tag{1}$$

For further details on Frisbee physics, we refer to the papers by Hubbard and Hummel [10] and Hummel [11], and for the actual implementation to our source code[1].

The game's virtual environment consists of a table in front of the player on which the Frisbee spawns after each throw as well as a board with information about the score, number of throws, misses and time tracking. 10 meters from the player is a round bullseye target with a diameter of 3 meters (see Figure 1). When a disc hits the target, it stays at the collision point as a reference, and the score is increased. The target is segmented into five rings with different score values assigned – the closer to the centre, the higher the score – to motivate the player to aim at the centre. For every throw, the disc's distance to the centre of the target was logged. In case it missed the target, we logged the distance and direction from the disc to the point where it intersected a plane extending the target's surface, or where it hit the ground if did not pass the target. This way, we not only tracked if the target was missed but also by how far and where it was missed.

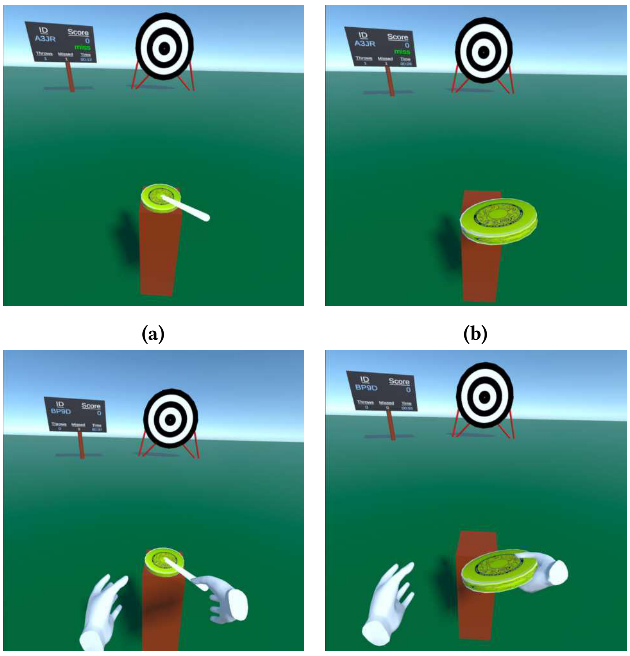
For the visualisation of the player's hands in one condition, we implemented floating hands that were animated to a gripping pose when picking up a disc and opened up again when releasing it. Participants in the other condition did not see their virtual hands. In both conditions, a white beam appears when pointing at the disc to indicate that it can be picked up. When the user then presses the hand trigger of the Touch Controller that lies under the middle finger, the disc is moved towards the hand and follows it. The interaction is displayed in Figure 2.

4 STUDY

We conducted a user study ($N = 29$) to observe the effect of hand visibility on the accuracy, presence, perceived body ownership and agency when throwing a Frisbee. Participants were asked to pick up and throw a disc onto a circular target with the goal of hitting as close to the centre as possible. We performed a double-blind study in which neither the participants nor the instructors of the study knew which condition had been selected for the specific trial at hand. The selection of the conditions of *no hands* or *hands* was counter-balanced across the sample. To avoid learning effects as a confounding factor on performance and habituation effects from the embodiment, we chose a between-group design, thus participants played the game only in one of the two conditions.

In total 29 people participated in the study: 15 without visible hands and 14 with visible hands. The age ranged from 20 to 71 with a mean of 33 ($SD = 14, 2$). In total, 16 participants identified as female and 13 as male. The condition without hands was tested by 6 female and 9 male participants, and the condition with hands by 10 female and 4 male participants. In both groups a small proportion of the participants was left-handed (2 out of 15 (13%) for the no hands condition and 3 out of 14 (21%) for the with hands condition), while the rest threw with their right hand. The self-assessment of how athletic they are and their proficiency in throwing a Frisbee also turned out to be balanced between the groups. Three of the participants had extensive prior experience in VR (2 in the group without hands, and 1 in the group with hands).

[1] https://github.com/MalteBorgwardt/VRisbee



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

Figure 2: The hand visualisations of the two conditions before and after grasping the Frisbee. From left to right: (a) Invisible hands before grasping the disc, (b) invisible hands holding the disc, (c) visible hands before grasping the disc, (d) visible hands holding the disc.

The study took place indoors in a spacious and closed room. Every player had 10 training shots and, subsequently, 30 trials to score as many points as possible in a standing position. The duration was unlimited and all participants were told explicitly they had as much time as they wanted or needed for their throws. The average time taken by the participants for their 30 trials was 3 minutes and 37 seconds. The whole experiment took on average approximately 40 minutes with the introduction, test trials, experimental trials, questionnaire, optional trials for fun, and interview. During the game, metrics for each throw were logged and stored on the VR headset after the experiment was finished. After completion, the participants filled in the following questionnaires on a laptop – all questions were shown in English with a German translation provided:

- Demographic questions: age, gender, height, handedness (left or right)
- Prior experience: athleticism in general, proficiency in throwing a Frisbee, experience with VR headsets
- Presence Questionnaire by Witmer & Singer [22] excluding the sound and haptic components
- Custom questions on body ownership as listed in Table 1.
- Custom questions on the control over the Frisbee as listed in Table 1.

Concluding the experiment, each of the participants was asked for feedback about their perception of the game and its mechanics.

Body ownership
I was aware of the location of my biological hands.
I was aware of the location of my virtual hands.
I felt that my biological hands were present in the virtual environment.
I knew exactly where my hands were in the virtual environment.
I was confused about my hands in the virtual environment. (reversed)
Control over the disc
How precisely were you able to control the direction of the throw?
Did the disc hit the spot you aimed for?
Did the disc behave the way you wanted it to?
How satisfied were you with the result you achieved?
How realistic was the throw?

Table 1: Custom questions of the questionnaire on body ownership and control over the Frisbee.

5 RESULTS

With the collected data we tested our hypotheses. We assume that hand visualisation has advantages, and therefore performed one-sided tests. In all our tests we used an alpha level of .05.

5.1 Accuracy

Since the data on **throwing accuracy** is not normally distributed according to the Shapiro-Wilk-Test ($p < .001$), we applied a Mann-Whitney U test. If we take all throws into account, we observe a significant difference ($p = .008$) with a negligible effect size of $r_{rb} = .097$. The mean accuracy was a distance to the centre of 1.672 m for the with hands condition and 1.958 m for the no hands condition. When looking solely at the throws not hitting the ground, the difference is not significant ($p = .096$). The respective means are 1.446 m for with hands and 1.474 m for no hands. This means, throws in the with hands condition were generally closer to the centre and therefore more likely to hit the target. When looking at the results of left-handed and right-handed participants separately, the scatter plots show an interesting characteristic, as illustrated in Figure 4. The hits and misses of right-handed throws tend to be deflected around an axis from bottom-left to top-right. For left-handed throws, the deflection tends to be from bottom-right to top-left, instead.

5.2 Questionnaires

To evaluate each subscale of the Presence Questionnaire by Witmer & Singer [22] and the two custom questionnaires on body ownership and control over the disc, we used Mann-Whitney U tests.

From the **Presence** Questionnaire, only the category 'realism' showed a significant difference ($p = .002$) with a large effect size ($r_{rb} = -.654$) in favor of the with hands condition. Box plots for 'realism' are shown in Figure 3. For the other categories 'possibility to act', 'quality of interface', 'possibility to examine', and 'self-evaluation of performance' there was no significant difference with p-values $p > .161$ for each category.

For the overall **body ownership**, the difference was not significant ($p = .051$). As this test is close to a significant result, it is conceivable that with a larger sample, a possible effect might be

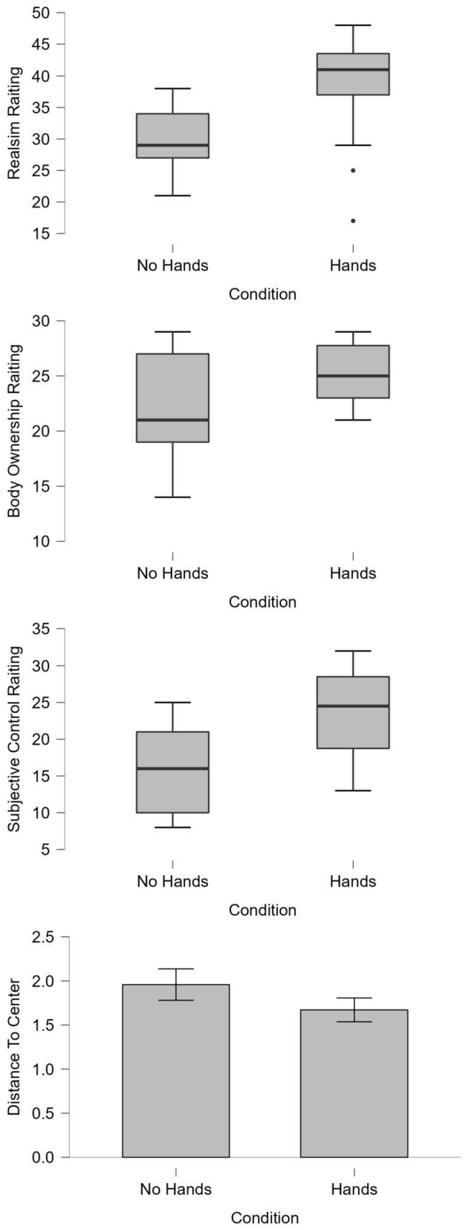


Figure 3: Box plots for the rating of realism, body ownership and subjective control and a bar plot for the accuracy with a 95% confidence interval for the two conditions no hands and hands.

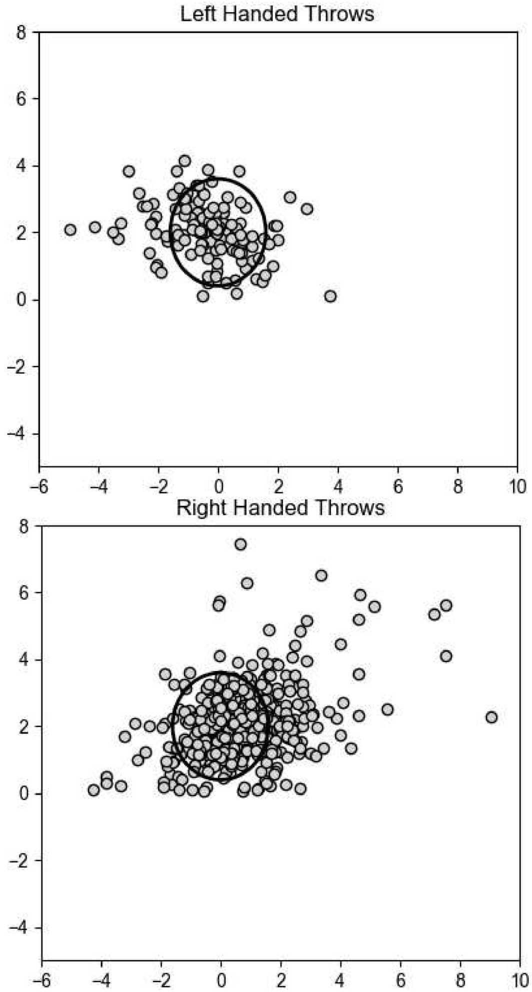


Figure 4: Scatter plots of the dispersion pattern with circles showing the actually circular target, top: left-handed throws, bottom: right-handed throws. The axis are displayed in meters.

measurable. In the distributions of the data visualised in Figure 3, we can see the distributions are close to each other but also have some differences. There is a large range of opinions for the without hands condition in contrast to the with hands condition.

The **subjective control** items show a significant difference (p = .004) with a large effect size of $r_{\text{rb}} = -.599$. The box plots for the data on subjective control are shown in Figure 3.

5.3 User Feedback

The overall feedback of the experiment was that the participants had fun. Many wanted to continue playing after the 30 throws and made 20-70 more throws. Participants reported that the controllers were not perceived as negatively influencing the throw, but three participants used the controller in different ways for picking up and throwing the virtual disc.

6 DISCUSSION

Our results show that hand visibility improves accuracy significantly but with a small effect size. Both groups have a similar data distribution, and the mean accuracy of the with-hands condition was slightly better. This only cautiously supports hypothesis H1 as the throwing accuracy does not strongly depend on whether the hands are visible or not. Since all participants interacted with the system for the first time and throwing in virtual reality generally is a challenging task, the accuracy of the throws might have suffered from unusual usage of the controllers. Still, our sample covered a wide range of ages and expertise in VR or sports, thus the result can be seen as representative.

A second look into the dispersion pattern shows two things. First, the scattering tends to be more prominent in the horizontal axis, which is in line with the findings of Zindulka et al. [23] as described in section 2, where overhand and underhand throws resulted in larger scattering around the vertical axis. Second, left-handed throwers would rather hit the left side of the target while the hits of right-handed throwers aimed more to the right. Given the movement of a natural backhand throw of a disc, this might indicate that the release timing of the disc with a controller is difficult and players tend to release the disc too late.

Regarding H2, we saw, that visible hands only had a positive influence on the presence subscale *Realism*. While visible hands induced a more realistic scenario, we could not identify other impacts on the feeling of "being there" in the virtual world, and being able to interact with it. This could be due to the fact, that even with invisible hands, participants were able to manipulate and throw the disc as they wanted to.

The results on body ownership showed no significant difference (H3) but are borderline. Therefore, we cannot assess if body ownership is affected by hand visibility or not. It would seem plausible considering the accuracy of the throws. With only a small impact through visible hands on the accuracy, this indicates that the visual focus while aiming lies mainly on the target and the launch is led by proprioception or muscle memory rather than visual cues.

In addition, with visible hands, the subjective control was rated significantly higher. The participants with visible hands had the feeling of having more control and being able to throw the disc more accurately (H4). This result could also be related to the fact that the system required the players to pick up a new disc manually, which could be a more difficult task with invisible hands, and therefore influence the feeling of control negatively.

6.1 Limitations

These findings are the subjective perception of the participants on their experience of the whole game. That includes not only the throw itself but also the mechanism of picking up a new disc to throw next. Here, players with a visible hand could have an advantage in hand-eye coordination, leading to a quicker feeling of success compared to the group with invisible hands. It is also important to mention that the disc does not perfectly behave the way it would in the real world. For example, the curve behaviour of the disc when it is thrown with an angle is not as prominent as with real discs, which could irritate participants with experience in throwing a disc.

6.2 Future Work

Future research in the area may include a wide range of options for throwing a disc in VR that can also help understand hand or body visibility. Those could be, for example, hand tracking for bare-hand interaction with no haptic feedback at all instead of using a controller. Alternatively, the Valve Index Controller could be used in this type of research as it allows tracking the position and pressure of every finger as well as releasing the grasp around the controller handle for releasing the held object. This might lead to a more realistic experience when grabbing and releasing a disc, and possibly a better outcome in terms of accuracy and precision. Another step towards realistic sensations can be haptic feedback gloves, which give users force feedback on all fingers, in addition to precise hand and finger tracking. Nevertheless, the weight and centre of mass of the virtual disc would not be rendered.

Further studies may also focus not only on the first few sets of throws but focus on long-term practice. This way, the influence of hand visibility can be investigated in more detail for experienced players as its effect might decrease with a more reliable muscle memory over time.

7 CONCLUSION

In this study, we analysed the influence of hand visibility on Frisbee throws. We found, that seeing virtual hands does not improve the accuracy of the throws considerably. Nonetheless, the subjective control over the disc was significantly higher in the group with visible hands. It seems, that visible hands raise the feeling of more control in the virtual world and support the user in secondary tasks such as picking up the disc before throwing it. Still, we could see similar results as Zindulka et al. [23] for throwing a virtual disc, in essence, that the release timing in general is a difficult challenge and has a strong impact on where the disc lands.

We also found that hand visibility has a positive significant influence on perceived realism, while the overall perceived presence did not benefit from visible hands. Also, we could not confirm that hand visibility has a significant positive influence on body ownership in disk throwing but inconclusive results. In contrast to the studies on kicking in VR by Bonfert et al. [3] and climbing in VR by Kosmalla et al. [14], our findings showed that a visual representation of the limbs alone does not suffice for a throwing task, which requires highly adjusted motor skills while already holding the virtual object.

ACKNOWLEDGMENTS

This work was partially funded by the Klaus Tschira Foundation, the KI-SIGS Project "AP 390" of the German Federal Ministry of Education and Research (BMBF), and the German Research Foundation (DFG) as part of the Collaborative Research Center (SFB) 1320 *EASE – Everyday Activity Science and Engineering*, University of Bremen (www.easecrc.org).

REFERENCES

[1] F. Argelaguet, L. Hoyet, M. Trico, and A. Lecuyer. 2016. The role of interaction in virtual embodiment: Effects of the virtual hand representation. In *2016 IEEE Virtual Reality (VR)*. 3–10. https://doi.org/10.1109/VR.2016.7504682

[2] Arnold Baca, Peter Dabnichki, Mario Heller, and Philipp Kornfeind. 2009. Ubiquitous computing in sports: A review and analysis. *Journal of sports sciences* 27 (09 2009), 1335–46. https://doi.org/10.1080/02640410903277427

[3] Michael Bonfert, Stella Lemke, Robert Porzel, and Rainer Malaka. 2022. Kicking in Virtual Reality: The Influence of Foot Visibility on the Shooting Experience and Accuracy. In *Proceedings of the 2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE Computer Society, Virtual. https://doi.org/10.1109/VR51125.2022.00092

[4] Karolis Butkus and Tautvydas Ceponis. 2019. Accuracy of throwing distance perception in virtual reality. In *IVUS*.

[5] Ryan Canales, Aline Normoyle, Yu Sun, Yuting Ye, Massimiliano Di Luca, and Sophie Jörg. 2019. Virtual Grasping Feedback and Virtual Hand Ownership. In *ACM Symposium on Applied Perception 2019* (Barcelona, Spain) *(SAP '19)*. Association for Computing Machinery, New York, NY, USA, Article 4, 9 pages. https://doi.org/10.1145/3343036.3343132

[6] Stefanie A. Drew, Madeline F. Awad, Jazlyn A. Armendariz, Bar Gabay, Isaiah J. Lachica, and Jacob W. Hinkel-Lipsker. 2020. The Trade-Off of Virtual Reality Training for Dart Throwing: A Facilitation of Perceptual-Motor Learning With a Detriment to Performance. *Frontiers in Sports and Active Living* 2 (2020). https://doi.org/10.3389/fspor.2020.00059

[7] Shaun Gallagher. 2000. Gallagher, S. 2000. Philosophical Conceptions of the Self: Implications for Cognitive Science. *Trends in cognitive sciences* 4 (02 2000), 14–21. https://doi.org/10.1016/S1364-6613(99)01417-5

[8] Jens Grubert, Lukas Witzani, Eyal Ofek, Michel Pahud, Matthias Kranz, and Per Kristensson. 2018. Effects of Hand Representations for Typing in Virtual Reality. 151–158. https://doi.org/10.1109/VR.2018.8446250

[9] David J Harris, Gavin Buckingham, Mark R Wilson, Jack Brookes, Faisal Mushtaq, Mark Mon-Williams, and Samuel J Vine. 2020. Exploring sensorimotor performance and user experience within a virtual reality golf putting simulator. *Virtual Reality* (2020), 1–8. Publisher: Springer.

[10] Mont Hubbard and S. Hummel. 2000. Simulation of Frisbee Flight. (01 2000).

[11] Sarah Ann Hummel. 2003. Frisbee Flight Simulation and Throw Biomechanics. (2003). https://morleyfielddgc.files.wordpress.com/2009/04/hummelthesis.pdf last access: 02.08.2022, Master Thesis, University of California, Davis.

[12] Konstantina Kilteni, Raphaela Groten, and Mel Slater. 2012. The Sense of Embodiment in Virtual Reality. *Presence Teleoperators & Virtual Environments* 21 (11 2012). https://doi.org/10.1162/PRES_a_00124

[13] Martin Kocur, Sarah Graf, and Valentin Schwind. 2020. The Impact of Missing Fingers in Virtual Reality. In *Proceedings of the 26th ACM Symposium on Virtual Reality Software and Technology* (Virtual Event, Canada) *(VRST '20)*. Association for Computing Machinery, New York, NY, USA, Article 4, 5 pages. https://doi.org/10.1145/3385956.3418973

[14] Felix Kosmalla, André Zenner, Corinna Tasch, Florian Daiber, and Antonio Krüger. 2020. The Importance of Virtual Hands and Feet for Virtual Reality Climbing. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3334480.3383067 event-place: Honolulu, HI, USA.

[15] Ke Ma and Bernhard Hommel. 2015. The role of agency for perceived ownership in the virtual hand illusion. *Consciousness and Cognition* 36 (2015), 277–288. https://doi.org/10.1016/j.concog.2015.07.008

[16] Seyyed Ahmad Mousavi, Mehdi Shahbazi, Elahe Arabameri, and Elham Shirzad. 2018. The Effect of Virtual Reality Training on Learning and Kinematics Characteristics of Dart Throwing. *International Journal of School Health* (2018).

[17] David Neumann, Robyn Moffitt, Patrick Thomas, Kylie Loveday, David Watling, Chantal Lombard, Simona Antonova, and Michael Tremeer. 2018. A systematic review of the application of interactive virtual reality to sport. *Virtual Reality* 22 (09 2018). https://doi.org/10.1007/s10055-017-0320-5

[18] Nicolas Nostadt, David A. Abbink, Oliver Christ, and Philipp Beckerle. 2020. Embodiment, Presence, and Their Intersections: Teleoperation and Beyond. *J. Hum.-Robot Interact.* 9, 4, Article 28 (may 2020), 19 pages. https://doi.org/10.1145/3389210

[19] Stefan Pastel, Chien-Hsi Chen, Katharina Petri, and Kerstin Witte. 2020. Effects of body visualization on performance in head-mounted display virtual reality. http://dx.doi.org/10.25673/36147

[20] Richard Skarbez, Frederick P. Brooks, Jr., and Mary C. Whitton. 2017. A Survey of Presence and Related Concepts. *Comput. Surveys* 50, 6 (Nov. 2017), 96:1–96:39. https://doi.org/10.1145/3134301

[21] Judith Tirp, Christina Steingröver, Nick Wattie, Joe Baker, and Jörg Schorer. 2015. Virtual realities as optimal learning environments in sport – A transfer study of virtual and real dart throwing. *Psychological Test and Assessment Modeling* 57 (01 2015), 57–69.

[22] Bob G. Witmer and Michael J. Singer. 1998. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments* 7, 3 (06 1998), 225–240. https://doi.org/10.1162/105474698565686 arXiv:https://direct.mit.edu/pvar/article-pdf/7/3/225/1836425/105474698565686.pdf

[23] Tim Zindulka, Miroslav Bachinski, and Jörg Müller. 2020. Performance and Experience of Throwing in Virtual Reality. 1–8. https://doi.org/10.1145/3313831.3376639

Publication P8

An Evaluation of Visual Embodiment for Voice Assistants on Smart Displays

Michael Bonfert, Nima Zargham, Florian Saade, Robert Porzel, and Rainer Malaka

This publication investigates the interaction of 60 users with a smart display with different agent representations. Three variants are compared in a between-group experiment: a disembodied agent (status quo), an artificial embodied agent, and a photorealistic embodied agent performed by a human actress. The quantitative metrics show no effect on the user experience. The rich qualitative analysis addresses the continuous on-screen presence of the agents, user considerations of agent appearance, and how the visualization influenced social aspects of the interaction.

My contribution: I made equal contributions to the research idea (*conceptualization*) and the study design (*methodology*) as NZ and FS. I contributed the majority of the quantitative statistical analysis (*data curation* and *formal analysis*). I wrote and edited the majority of the manuscript (*writing – draft, review, & editing*) and created all figures and presentation material (*visualization*). I supervised the project together with NZ (*supervision*).

Published in the *Proceedings of the 2021 ACM Conference on Conversational User Interfaces (CUI '21)*

An Evaluation of Visual Embodiment for Voice Assistants on Smart Displays

Michael Bonfert*
bonfert@uni-bremen.de
Digital Media Lab
University of Bremen, Germany

Nima Zargham*
zargham@uni-bremen.de
Digital Media Lab
University of Bremen, Germany

Florian Saade
florian.saade@uni-bremen.de
University of Bremen, Germany

Robert Porzel
porzel@uni-bremen.de
Digital Media Lab
University of Bremen, Germany

Rainer Malaka
malaka@uni-bremen.de
Digital Media Lab
University of Bremen, Germany

Figure 1: Our prototype of a smart display with an embodied voice assistant agent performed by an actress

ABSTRACT

Smart displays augment the concept of a smart home speaker with a touchscreen. Although the visual modality is added in this device variant, the virtual agent is still only represented through auditory output and remains invisible in most current products. We present an empirical study on the interaction of users with a smart display on which the agent is embodied with a humanoid representation. Three different conditions are compared in a between-group experiment: no agent embodiment, a digitally rendered character, and a photorealistic representation performed by a human actress. Our quantitative data do not indicate that agent visualization on a smart display affects the user experience significantly. On the other hand, our qualitative findings revealed differentiated perspectives by the users. We discuss potentials and challenges of embodying agents on smart displays, reflect on their continuous on-screen presence, present user considerations on their appearance, and how the visualization influenced the politeness of the users.

*Both authors contributed equally to this research.

CUI '21, July 27–29, 2021, Bilbao (online), Spain

ACM ISBN 978-1-4503-8998-3/21/07...$15.00
https://doi.org/10.1145/3469595.3469611

CCS CONCEPTS

• **Human-centered computing → Natural language interfaces**; User studies.

KEYWORDS

Voice Assistants, Conversational Agents, Embodiment, Smart Displays

ACM Reference Format:
Michael Bonfert, Nima Zargham, Florian Saade, Robert Porzel, and Rainer Malaka. 2021. An Evaluation of Visual Embodiment for Voice Assistants on Smart Displays. In *3rd Conference on Conversational User Interfaces (CUI '21), July 27–29, 2021, Bilbao (online), Spain.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3469595.3469611

1 INTRODUCTION

The use of voice interaction is spreading widely. Current voice assistants (VA) have broad capabilities in helping the users for different purposes, such as smart home control, work management, scheduling, gathering information, navigation, communication, education, or entertainment. Voice user interfaces are available in mobile phones, personal computers, cars, smart speakers, and other devices to make the interaction easier, more accessible, and more natural. The affordances and accessibility facilities of home assistants – also referred to as smart speakers – differ from those of VAs on smartphones. Interaction with home assistants is possible from a distance and enables to control smart home appliances [46]. With the device's exterior, the system has a physical embodiment within the room, however, not the assisting agent.

Research suggests that emulating human qualities affects how users feel towards VAs [11]. The experiences can be different between users depending on their own personalities [14], but also depends on the assistant's personality. Currently, the personality of a VA is primarily conveyed by its voice, linguistic characteristics of its answers, designated personifications as its name (e.g., Alexa instead of the product name Amazon Echo), and its physical device design [4, 5, 48]. Moreover, research has shown that the identified gender of an agent has an impact on the user experiences [7]. The visual presence of current smart speakers is limited to the device's casing and abstract animations or LEDs that illustrate the assistant's state or audio output. To visually convey personality and human characteristics, the development of virtual assistants could focus further on embodiment. Researchers have studied diverse types of embodiment for conversational agents [1, 24, 31] as well as their visual attractiveness [25]. Embodied virtual agents have become a natural extension of conversational interfaces by enriching the experience visually [2, 8, 43, 56].

A novel opportunity to embody VA agents are smart displays. This new product category of home assistants is equipped with a screen for visual output and touch input. Prominent examples from the consumer market are the Amazon Echo Show and the Google Nest Hub. These devices complement the features of a voice assistant with the possibilities to, for instance, look at pictures, watch videos, browse recipes, or display the smart front door camera.

Moreover, we see the potential in the screen to enhance the visual presence of the virtual agent. Therefore, we conducted a study on the user experience (UX) during the interaction with smart displays featuring an embodied agent. In this research, we pursue the following two research questions:

RQ1: How does the user experience change if a voice assistant agent is visually embodied on a smart display?

RQ2: How does the degree of visual realism of the embodied agent influence the user experience?

Building on prior research on attractiveness, gender, and appearance of embodied agents, we contribute an investigation on displaying an embodied VA agent and, moreover, its visual realism. This is done for the novel use case of a smart display considering social implications of the continuous agent presence in the room. We present an empirical study exploring user interactions when engaging with one of three different smart display prototypes: one with a disembodied agent, one with an artificial, digitally rendered embodiment, and one with a photorealistic embodied agent performed by a human actress. Our quantitative analysis includes two standardized UX questionnaires and the expressed politeness during the interaction. In semi-structured interviews, we collected further impressions, preferences and expectations by the participants.

2 RELATED WORK

In this section, we discuss research on agent embodiment with focus on voice assistants, the Uncanny Valley and gender implications.

2.1 Embodied Agents

Embodied conversational agents are computer-controlled characters that can interact with people using natural language and engage in a dialog [9]. They can use facial expressions, gestures, and eye gaze to enable natural, multimodal human-computer communication. Numerous studies have explored how embodiment and its different forms, as well as a lack of a body, can influence human-machine interaction and users' trust [13, 17, 20, 49] and engagement [26]. One of the most controversial examples of a virtual assistant with a visual embodiment is *Clippy*, an animated paper clip appearing in Microsoft Office 97. It was not well received amongst users and it failed to deliver on the promise of interface agents [57]. Research has found that using humanoid embodiment and voice influences users' perceptions of social presence [3, 47]. This presence of an agent can affect the relationship with a user in many aspects, such as trust and respect [3, 21].

Users treat the system more like a person when an agent has an embodiment [30, 55]. Castillo et al. believe that state-of-the-art embodied conversational agents can change their perceived personality through appearance and behavior [10]. An embodied agent can leverage various means of non-verbal communication to better engage with users beyond speech [56]. Previous work suggests that users' perception of an embodied VA's personality is not just dependent on its visual or audible output. Researchers believe that personality is experienced in a multimodal manner and if designers only focus on either voice or facial characteristics to design personality, they will most probably not succeed [10].

2.2 Voice Assistant Embodiment Across Applications

People feel higher levels of social presence when there is a visual representation available, as the comprehensive review on social presence literature by Oh et al. shows [44]. Hernández-Trapote et al. found that users who interacted with an embodied agent had greater privacy concerns but also perceived the interaction as more pleasant compared to using a voice-only interface. In their study, the authors found no significant difference in user preference [20]. In contrast to avatars depicting a specific person, an embodied agent can be designed in any conceivable way depending on the given context and purpose. Wang et al. conducted a study on interactions with virtual agents in augmented reality. They compared four agent representations: voice-only, non-humanoid, full-size humanoid, and miniature humanoid. The experiment showed that both humanoid and non-humanoid agents were acceptable for users. However, having an agent visualized as a smart speaker strongly impacted users' conception of the agent not being human – even more than without visualization [56]. In virtual reality (VR) environments, Schmidt et al. showed major benefits for both embodied and thematically related audio-visual agent representations which positively affected the overall user experience in the context of a VR exhibition space. They also found that agent embodiment induces a higher sense of spatial and social presence [50].

With the aim to support information workers to be more productive and focused, Grover et al. designed and compared two productivity agents: a text-based agent, similar to a chatbot, and a virtual agent with a video embodiment. Their results show that users felt more productive and less distracted when being assisted by the embodied agent [16]. These findings are in line with a recent investigations on the effects of VA embodiment in augmented reality (AR). Kim et al. found that users performed better in collaborative decision making when interacting with a VA and reported a significantly lower task load when it was embodied [28]. Kim et al. further observed that users perceived agents in AR as more aware of and able to influence the real world if they are embodied [27]. Similar research has been done in virtual reality environments [50, 51].

The influence of human-like agent behavior was the focus of a study by Mayer et al. who assessed multimedia learning when being taught by on-screen agents. The team measured better performance in learning and recalling information when the agent behaved more like a human in speech and gestures [37]. The attractiveness of virtual agents has also been a topic of research. In a study by Khan and De Angeli, the users formed and maintained a better evaluation of attractive agents independent of the interaction with the agent [25]. It has been demonstrated that an agent's attractiveness may be even more important than its reliability [58].

2.3 Gender Implications

Researchers have extensively expressed their concerns on gendered agents as it can easily reproduce a stereotypical gender script [12, 54, 59]. Most of the common voice assistants available in the market set a female voice as default in most countries, which can amplify gender stereotypes [22]. A study by Nass et al. [42] suggests that even computers with minimized gender cues in the voice output evoke gender-based stereotypical responses. Authors tested three gender-based stereotypes without any gender indicators but vocal cues and witnessed stereotypes in all cases. In another study, Nass and Moon showed that users prefer to hear praises from a male agent rather than the same comments from a female agent [41]. Hwang et al. [22] categorized three distinct characteristics of bodily display, subordinate attitude, and sexualization to investigate the reflections of gender stereotypes toward women in female-voiced VAs. The authors suggest that such stereotypical traits could create a power dynamic between users and female agents. The described studies provide insights into the application of embodied agents across different mediums, use cases and characteristics. Our work extends research on embodied conversational agents to the domain of smart homes by bringing visualizations of a voice assistant to smart displays. Considering the large design space of possible agent visualizations, the question arises how close to a human appearance these should be. Thus, the investigation considers the degree of visual fidelity of the embodiment.

2.4 The Uncanny Valley

The term "uncanny valley" refers to a person's adverse reaction to robots that look and behave almost like a human, but not quite [39]. This effect has furthermore been investigated with any type of human-like entity or object, such as dolls, masks, facial caricatures, movie characters, avatars, and embodied agents [53]. Studies indicate that realistic humanoids can be appealing [18, 35, 38], but to achieve this, a number of aspects need to be considered. The artificial humanoid must attain a certain level of integrated social responsiveness and aesthetic refinement to appeal to the users [18]. Previous research has established that the uncanny valley effect emerges when there are abnormal features, or an insufficient degree of realism [53].

Some studies have explored the uncanny valley hypothesis in terms of human avatars [35, 38]. MacDorman et al. believe that a computer-generated face is not necessarily eeriest when it looks nearly human and argue that even abstract faces can look uncanny [35]. Guidelines for virtual character design by Schwind et al. recommend consistency in realism and deliberate stylization to avoid uncanniness [52]. To avoid uncanny valley effects in our study while still comparing cartoony to highly realistic embodiments of a VA agent, we decided to have an actress perform the agent for the photorealism condition. For most practical applications, this is obviously not an ecologically feasible solution, but provides clearer results in the context of this study.

3 PROTOTYPE DESIGN

We designed three versions of a smart display for the purpose of this experiment: one with a disembodied agent (DEA), one with a digitally rendered, artificial embodied agent (AEA), and one with a prerecorded, photorealistic embodied agent (PEA). All versions had the same functionality and only differed in appearance. We chose a female agent to reflect the predominance of female assistants in current consumer products with the intention to avoid a novelty bias [23]. The VA was called "Joy" and spoke the local official language German.

Disembodiment Agent (DEA) | This version was designed to resemble the current status quo of smart displays with no agent embodiment. The users would only hear the agent's synthetic voice. We generated the voice with the online Text-To-Speech (TTS) tool Natural Readers[1].

Artificial Embodied Agent (AEA) | For this version, we created a digitally rendered, animated visualization to represent the agent on the smart display. It shows a female, about 30-year-old character with blonde hair, light-colored skin and a dark blue dress as can be seen in Figure 2. The appearance reminds of a news anchor in the style of *The Sims*. We compared a variety of available options in an informal pre-study and found this character as best corresponding to the selected voice. To create the renderings, an actor performed in front of a webcam as input for FaceRig[2] to animate the virtual character. The video output was merged with the same TTS voice used for the DEA condition with synchronized lip movements.

[1] https://www.naturalreaders.com
[2] https://facerig.com

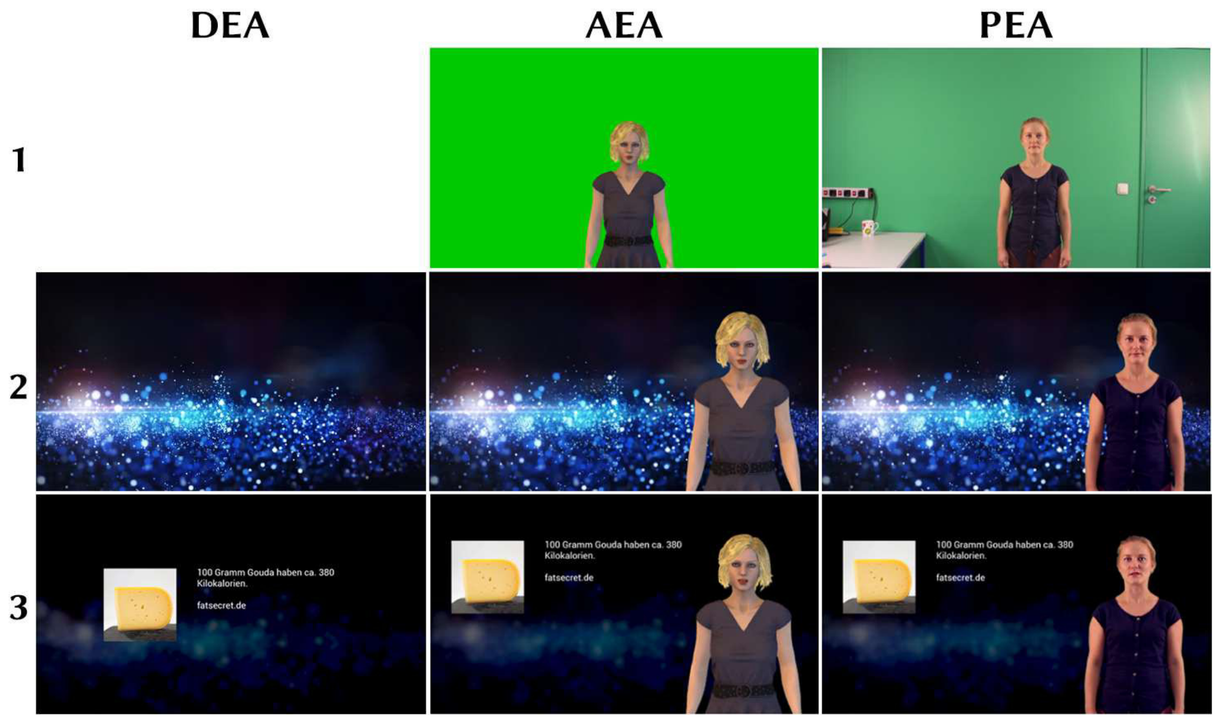


Figure 2: The three stages of the video creation process for the three conditions: 1) Recording or rendering with a green screen, 2) Replacing the background, 3) Augmenting with information cards. The second stage is used as idle loop.

Photorealistic Embodied Agent (PEA) | For this prototype version with a highly realistic embodiment, a theater actress was recorded. She was instructed to perform as similar as possible to the artificial character in terms of intonation, facial expressions, and body language. We refrained from using the TTS audio on account of lip synchronicity and to avoid a mismatch of visual and auditory coherence. The actress and her clothing were selected to resemble the AEA visually. All utterances were recorded in front of a green screen to be used as an overlay for the content.

3.1 Prototype Implementation

For the video and audio output of the smart display, we prepared media snippets of all responses needed for the experiment execution in each condition. Each snippet consisted of a dark, dynamic background, an information card, the audio track, and – where applicable – an agent embodiment, as illustrated in Figure 2. The information cards contained text and images related to the user's commands. They appeared when the assistant initiated the response and faded out when the task was performed. Between tasks, a dynamic idle video was looping. The smoothness of the transitions depended on the timing of the next inquiry, which affected all conditions equally. For the AEA and PEA conditions, the agent embodiment was added as an overlay on the bottom right without overlapping the information cards. The screen layout in the DEA condition was centered to avoid empty space where the agent would be shown in the other versions.

To ensure reliable system operability, we used a Wizard of Oz approach in this study. The Wizard sat in an adjacent room and controlled the smart display. This was disclosed to the participants after the study. The technical setup is illustrated in Figure 3. The prototype was assembled from a Nexus 7 tablet and a Bluetooth speaker. For mounting the components in a way to appear as a smart speaker, three tailored parts were manufactured with 3D printing and laser-cut acrylic glass. The Wizard listened to the user's commands via Skype which was running silently on the tablet in the background. The responses were triggered with the help of a structured playlist on VLC media player to provide an instantaneous responsiveness of the system. Via Splashtop, the video on the Wizard's laptop was streamed to the tablet. Until the user continued with the next inquiry, the system looped an idle sequence that continuously showed the agent. For the analysis of the user's language, an audio device within the room recorded the experiment. For realizing one of the experiment tasks concerning smart home appliances, we used a smart light bulb by Philips Hue activated with a remote control by the Wizard.

4 EXPERIMENT

We evaluated our prototypes in a Wizard-of-Oz experiment with a between-groups design in which the participants ($N = 60$) interacted with one of the prototypes to complete a specified set of tasks. The condition assignment was pseudo-randomized between three equally distributed groups of 20 users each.

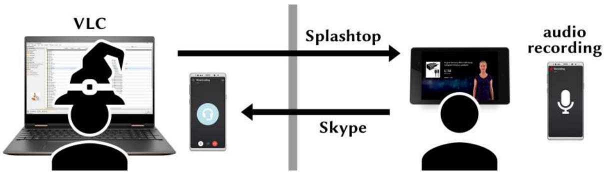


Figure 3: Technical setup of the experiment: The Wizard listens to the user's commands via Skype and triggers the appropriate video snippet in VLC media player which is transferred to the smart display via Splashtop. An audio device records the experiment.

4.1 Participants

We recruited 60 participants for this experiment (47 male, 13 female), between 14 and 38 years of age (M = 24.0, SD = 5.5). About half of the participants were students and 42% had a computer science background. All participants owned a smartphone. The majority of our participants (71.7%) stated that they rarely use a voice assistant on their phone. 11.7% never used a VA once. The other 16.7% use it at least several times a week. Concerning VAs in smart homes, 16.7% indicated using smart speakers regularly. The groups of phone VA users and smart speaker users have an overlap but are not identical. Only one participant had prior experience with a smart display and uses it daily. We conducted the experiment in the local official language German to avoid language barriers.

4.2 Procedure

After giving informed consent, all participants filled in a questionnaire about demographics and prior VA experience. Afterwards, the participants watched a short tutorial video on a separate screen outlining eleven predefined tasks to perform. We provided the participants with a paper list of the tasks to accomplish. Then, the test began with the smart display showing the idle video sequence, including the embodied agent if applicable. The first interaction was initiated by the user.

The activities represent a morning scenario and were designed to include a broad range of everyday commands following an analysis of typical home assistant usage [29]. These included, for example, turning on the light, playing music, retrieving information, setting a timer, or ordering a product online. All tasks are listed in the Appendix A. Sometimes, the participants forgot to use the wake word yielding in no reaction of the VA. When the user asked questions that were not included in the command list, the system explained that it cannot help with this. To ensure comparable interaction experiences and levels of frustration, one simulated failure to comply was included in each session even when a participant followed the task list strictly.

After finishing the tasks, the participants filled in a paper questionnaire comprising the User Experience Questionnaire (UEQ) [32] and the AttrakDiff Short Questionnaire [19]. Both scales are validated and established measurement instruments with a similar underlying theoretical construct to assess the pragmatic and hedonic qualities as well as the attractiveness of a system. The questionnaires provide an authoritative, quantitative measure of the user's subjective experience. In combination, the collected data can be compared to confirm the reliability of the measurements. Finally, the experimenter conducted a brief semi-structured interview covering aspects of reliability, trust, agent appearance, individual preferences, and permanent on-screen presence. At the end, all participants were demonstrated the alternative system versions to allow a comparison, despite the between-groups approach. This was done last in the interviews to not influence any prior assessments. Everyone was shown the same, complete sample snippet from both unfamiliar conditions to ensure comparability. The experiment and interview were recorded acoustically for later analysis. Each test session took 30 − 50 minutes.

4.3 Data Analysis

Two participants gave contradictory answers within three or more scales of the UEQ. As recommended by the handbook, their ratings were excluded from the analysis as it can indicate random or not serious answers [32]. Further, one participant did not fill in the AttrakDiff. For both questionnaires, the visual interpretation of the histograms raised doubts about the normal distribution of the data. This assumption was supported by Shapiro-Wilk tests. Therefore, we applied non-parametric tests. We ran Kruskal-Wallis tests to check for group differences between the three conditions. Due to technical issues, only $n_{quant} = 50$ audio recordings of the experiment sessions were complete and valid for statistical analysis. The unequal distribution between the conditions (DEA: 19, AEA: 16, PEA: 15) was considered for the statistics. The number of "Thank you" and "Please" utterances per user was compared with Mann-Whitney U tests between the groups. For all statistical tests, we applied an alpha level of .05.

Regarding the qualitative data, three interviews could not be analyzed due to data loss from a defective SD card. The other $n_{qual} = 57$ interview recordings were systematically examined (DEA: 20, AEA: 18, PEA: 19). Three researchers agreed on a coding system that was generated from a random selection of ten interviews. Then, all recordings were analyzed, coded along this categorization, and summarized. Additionally, we collected insightful and unique statements.

5 RESULTS

We present our findings in three sections: quantitative system evaluation, suggestions for the visual appearance, and considerations regarding the permanent presence of the agent.

5.1 Quantitative System Evaluation

From the standardized questionnaires, we learn that all three conditions, with a disembodied agent (DEA), with an artificial embodied agent (AEA), and with a photorealistic embodied agent (PEA), can result in comparably good user experiences. For all groups, the User Experience Questionnaire, rated from −3 to +3, shows overall high ratings for *attractiveness* ($Mean_{DEA}$ = 1.38 ± Standard Deviation$_{DEA}$ = .61; M_{AEA} = 1.12 ± .92; M_{PEA} = 1.58 ± .81) and the *pragmatic qualities* (M_{DEA} = 1.67. ± .59; M_{AEA} = 1.43 ± .68; M_{PEA} = 1.71±.46). The scores of the *hedonic qualities* (M_{DEA} = .84± .72; M_{AEA} = .81 ± .94; M_{PEA} = .99 ± 1.11) are below average according to the UEQ handbook. The data distribution of the single subscales yielding in these aggregated scores are illustrated in the

Table 1: Statistics on the UEQ and AttrakDiff analysis with Kruskal-Wallis H and asymptotic significance p for the subscales Attractiveness, Pragmatic Qualities, and Hedonic Qualities.

UEQ	**Attr.**	Perspicuity	Efficiency	Dependability	**Prag. Q.**	Stimulation	Novelty	**Hed. Q**
H	3.585	1.721	3.023	1.268	1.200	0.306	2.954	1.358
p	.167	.423	.221	.530	.549	.858	.228	.507
AttrakDiff	**Attr.**	**Pragmatic Qualities**				**Hedonic Qualities**		
H	1.131	1.888				0.437		
p	.568	.389				.804		

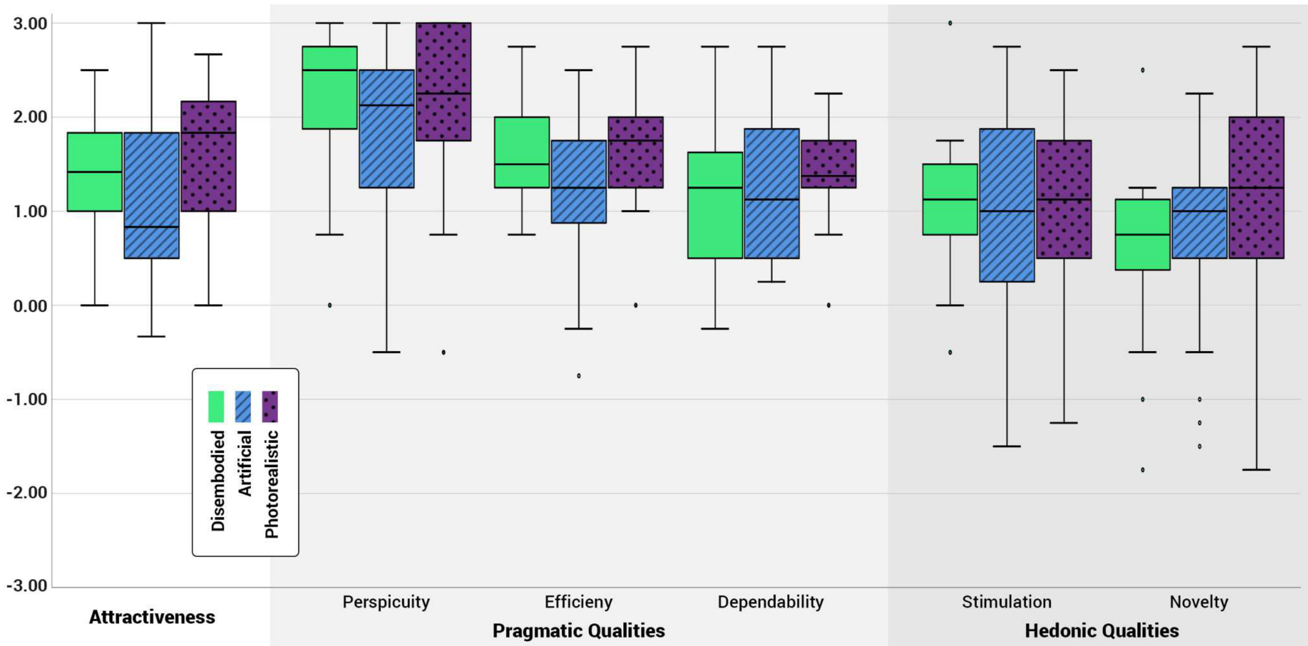


Figure 4: A box plot showing the distribution of ratings along the six subscales of the User Experience Questionnaire (UEQ) comparing the three conditions *Disembodiment Agent* (green), *Artificial Embodied Agent* (blue), and *Photorealistic Embodied Agent* (purple)

box plot in Figure 4. The data show no significant differences between the conditions on any of the subscales or aggregated scores ($p > .05$) as shown in Table 1.

These measurements match the user ratings on the AttrakDiff Short Questionnaire. In line with the UEQ results, no group differences were observed ($p > .05$). This measurement tool classified all our tested systems clearly as "task-oriented" due to high ratings for *pragmatic qualities* ($M_{DEA} = 1.63 \pm .80$; $M_{AEA} = 1.49 \pm .72$; $M_{PEA} = 1.78 \pm .83$) and medium ratings for *hedonic qualities* ($M_{DEA} = .79 \pm .73$; $M_{AEA} = .69 \pm .76$; $M_{PEA} = .75 \pm 1.12$). Like the UEQ, the AttrakDiff evaluates the system's *attractiveness* and yields a similar outcome with medium to high ratings ($M_{DEA} = 1.48 \pm .83$; $M_{AEA} = 1.15 \pm .90$; $M_{PEA} = 1.24 \pm 1.11$).

After the experiment, we showed the participants how the other two system versions look like and asked them to choose their preferred version. More than half decided for the photorealistic agent (51.8%). Only one out of eight users would select the artificial agent (12.5%) and every third person favored the version with a disembodied agent (35.7%). We found a bias in preference for the system version that the user was familiar with from the experiment, especially pronounced for the AEA condition: 71% of the people who favored the artificial embodiment in the interview used it earlier in the study (expected value: 33%). Only two people from another condition preferred the artificial variant. Moreover, 83% of the users who worked with the photorealistic agent embodiment preferred this version. Similarly, 60% of the participants from the DEA condition preferred to have no agent embodiment.

From recordings, we analyzed the users' verbal input in terms of expressed courtesy towards the VA. Overall, 28% of the users said "please" in at least one of their inquiries with no significant differences between the conditions. However, while 47% thanked the disembodied agent and 44% said "thank you" or "thanks" to the artificial agent, only 13% expressed thankfulness toward the photo-realistic agent. With an average of $M = 0.13$ thankful utterances per session, participants in the PEA condition used significantly less thankfulness indicators than the DEA users with $M = 0.95$ ($U = 91.0$, $p = .030$, $\eta^2 = .094$). The difference in comparison to the AEA users ($M = 0.75$) is borderline but not statistically significant ($U = 80.5$, $p = .050$, $\eta^2 = .079$).

5.2 Agent Appearance

When we asked our participants about the ideal appearance of the agent, we observed attitudes that can be roughly categorized as pragmatic, personal, and playful. Firstly, the pragmatic group argued for omitting any agent visualization as there is no functional purpose of it in voice interaction. It was described as an unnecessary distraction taking up space, that could be used to display important content. If the agent was supposed to be embodied, users in this group would typically prefer a non-humanoid appearance. As abstract representations, they proposed eyes, an emoji, or minimalist animations such as waves, dots, or a point cloud. A user argued that it should be visually clear that the interlocutor is a machine and not a human. For this, a robot was suggested.

Secondly, the group with a preference for a more personal interaction were in favor of a human-like embodiment. A typical reason for this was that it is perceived as more trustworthy and more natural. However, three users were concerned about the authenticity of a digitally rendered visualization. It was perceived as "creepy" as it looked "not human enough" (P23), for example due to the lack of gestures. Seven participants were in support of a cartoon style. Three participants would like to be assisted by "an attractive woman" and one even specified the preferred hair color. For almost half of our participants (45%), the gender of the agent does not matter. Most of the rest would rather have a female (42%) than a male agent (13%). Several users (22 of 57=39%) explained that the agent should ideally be of similar age as themselves. It should not seem "too young, so it is reliable" (P34) and knowledgeable. Others were concerned about the agent being much older than themselves because it might make them feel parented or patronized.

Thirdly, the playful users proposed creative and fun ideas for the agent embodiment. These included animals, dinosaurs, and fantasy creatures. 13 participants requested celebrities, such as musicians, actors, or athletes, but also fictional characters from pop culture, such as Spiderman, Darth Vader, Pokémon, Hermione, Dobby, Rick and Morty, Yoshi, or – as "someone who fits into this role" (P35) – Batman's butler Alfred. Even a modern adaptation of Microsoft Office's *Clippy* was proposed. One user suggested to show the user's self-avatar as the agent. Moreover, someone proposed changing characters for specialized task areas, e.g., a depiction of a grandmother for recipes. One person advocated gender-neutral solutions to not further increase the bias in the perception of children, that the typical assistant should be female.

5.3 Permanent Presence of the Agent

Several users (12 of 57=21%) appreciated that the agent was always visible – also in the idle state. It was perceived as steady availability of the system: while the agent is present it can obviously be addressed at any time. In contrast, the majority (63%) of the users in our sample expressed that they would like the assistant to disappear from the screen after a task was performed and only reappear when called upon. Most often, this was explained with the awkward feeling of being watched by the agent. One user described the impression that "the device is alive" when there is an agent starring at him (P45). Nine users found it unsettling that the assistant seems to be waiting for them: "it feels like [the agent] expects something" (P33). Another participant was concerned that "when there is a human [agent] idling around, it would be very creepy" (P38). Six participants would appreciate the transition as an indicator for the successful recognition of the activation word. By some users, the agent's unchanged presence was misinterpreted as a permanent responsiveness. This led to misunderstandings in which the users continually omitted the activation word and were frustrated by the lack of feedback.

Five users (9%), who prefer the agent to disappear when idle, speculated about the design of the transition. For P12, it is important to avoid a sudden disappearance because in reality, people do not suddenly vanish. Similarly, one participant proposed a reality-inspired design in which the agent would walk in and out of the frame as needed. User P37 suggested a humorous adaption of this idea. She would like if the agent occasionally walked through the screen as when passing by, or read a newspaper while not needed.

6 DISCUSSION

In this study, we set out to understand the UX during the interaction with different visual representations of a smart display agent (RQ2) and compared it to a system with a disembodied agent (RQ1). With the two standardized UX questionnaires (UEQ and AttrakDiff Short), no significant differences between the conditions were found in terms of pragmatic qualities, hedonic qualities, or attractiveness of the systems.

Considering the qualitative findings, however, it is evident that an embodied agent does influence the interaction in various ways, beyond the measurements of the standardized instruments that we applied. The discrepancy between quantitative and qualitative results might be due to the broad range of UX aspects that the universally applicable questionnaires cover – which were found to be similar in all conditions during the short-term usage in our lab experiment – while the insights from the interviews mostly concerned social context, the imagined usage in a home environment, as well as design speculations specific to smart displays. These findings could hardly be brought to light with standardized scales but provide exciting avenues for future research. In the following, we will discuss what aspects are promising for a future, more targeted quantitative examination.

6.1 Embodied vs. Disembodied Agent

A third of our sample (35.7%) would prefer to use the status quo of a smart display with no depiction of the agent. Reasons for not showing a visualization were mostly of pragmatic nature. For a voice user interface, it was regarded as unnecessary, distracting, and blocking space that could be used for more relevant content. However, the pragmatic qualities of all systems yielded similar ratings and did not reveal advantages of not displaying an agent regarding how efficient, clear, fast, or predictable the system was perceived.

Every second participant (51.8%) preferred the version with a photorealistic agent. Although we observed a tendency towards preferring the system familiar from the experiment, especially noticeable in the AEA condition, only one out of eight (12.5%) users favored the artificial visualization. This is in line with results by Hernández-Trapote et al. [20] who assume a "balance of likeability and rejection factors" causing ambivalent preferences concerning agent embodiment. Similarly, our results demonstrate that embodiment is not beneficial, in principle, but depends on its implementation.

One of the key advantages of conversing with an embodied agent was explained with its higher subjective trustworthiness, supporting the findings of previous works [17, 49]. Similarly, in accordance with existing research [30, 55], the interaction was perceived as more natural with an embodiment. The possibility to see the agent seems to make it more approachable and dependable compared to only hearing the volatile voice. While for some users it was only important to visually focus on the interlocutor independent of its appearance, others had clear ideas of how it should look like. A group of users explained that it should be obvious that they are talking to a machine, so they are not led to believe that they are speaking with a human. For this, abstract and non-humanoid representations were suggested. Further, several playful and fun concepts were shared by the participants, including fictional characters, animals, or mythical creatures.

Overall, we learned a wide variety of conflicting preferences and reasons and can therefore assume that no universal solution will satisfy the expectations of all users equally. Consequently, we recommend enabling smart display users to determine whether an embodied agent is displayed and to customize its appearance to their liking.

6.2 Humanoid Appearance of the Agent

Among the participants who like the idea of seeing the agent, only a fifth preferred the prototype version with artificial, cartoon-style rendering. The users were skeptical about the artificial embodied agent as it was described as "not human enough" indicating an uncanny valley effect [40]. Indeed, the visualization in our AEA condition was technically not sophisticated, for example, due to the lack of gestures or detailed micromotions which influence the perceived humanness [37]. On the other hand, some users liked the deliberate cartoony realization, because the humanoid shape conveys a human-like conversation style, while the style maintains the obvious artificiality of the interlocutor.

80% of the users, who were in favor of displaying the agent, liked the photorealistic embodied agent the most. This is a notable outcome considering the mismatch of visual realism and behavioral artificiality. As a meta-analysis shows, the literature describes various differences in the perceived social influence of human-controlled avatars compared to computer-controlled agents independent of their degree of visual realism [15]. A factor that might have effected the participants' preference for a specific version might have been the different voices used in the conditions. We used a computer-generated voice using a TTS tool for the DEA and AEA conditions, and recorded a human voice for the PEA condition. This design decision was made to keep the system variants as coherent as possible in terms of audio-visual match and synchronicity. This consistently conveys to the user that one version is entirely artificial and the other as realistic as possible at the cost of using different voices.

Of course, recording actors to embody agents that are meant for universal application is not feasible outside a Wizard of Oz experiment. However, our results clearly indicate a preference of life-like realism over uncanny renderings or cartoony stylization. We, therefore, advocate for sophisticated, photorealistic renderings or alternatively fully abstract visualizations, depending on the requirements, target group, and objectives of the system. Outside of professional context, entertaining approaches with funny and fictional characters can be considered as an additional option for consumer products. Offering famous characters from pop culture could appeal to fans and create a fun experience.

Nearly half of our participants reported that the agent's gender is not important for them. Three quarters of users, who stated a preference, prefer to talk to a female agent which supports previous literature on gender stereotypes with conversational agents [7, 22]. This could be explained with habituation as participants mentioned that they are used to female VAs, which is still the default setting in most popular consumer products. Another explanation could be the sample skew toward male participants (78.3%). We also observed a sexualized component in the relationship to the agent, as a few participants wanted to have an attractive agent with customizable appearance, e.g., preferred hair color. These findings align with previous research by Khan and De Angeli regarding the attractiveness of embodied agents [25].

With an already pronounced gender bias in the VA market and a clear status imbalance in the interaction, we advocate for gender-aware solutions. While gender-ambiguous voices could be an apparent solution, Sutton argues that also other factors than voice can lead to binary assumptions on an agent's gender making the voice ambiguity redundant [54]. This becomes especially evident for embodied VAs and requires careful considerations for gender-sensitive interfaces. An alternative approach could be a balanced team of agents with various genders co-embodying the smart display [34].

Our findings showed that the assumed age of the agent could affect the user's assessment of reliability, which corresponds with findings by Marin et al. [36]. According to the participants, the agent should appear experienced, hence should not be too young. On the other hand, younger users explained that they do not want the agent to look much older than themselves since they could feel patronized or mothered. Participants described the ideal age as similar to their own.

6.3 Social Aspects and Awkward Presence

We observed that the participants expressed significantly less thankfulness towards the PEA compared to the DEA and AEA. This finding might seem counter-intuitive as we could expect more courtesy in interactions with a more realistic assistant. Contrasting views among users have been observed whether a voice assistant is entitled to politeness [6]. Another explanation could be the perceived real-time processing of the system. The reactions by the actress must have been recorded before the interaction; hence, the agent cannot rejoice in the expressed thankfulness. The user might assume a predefined emotional state that cannot be influenced. Whereas, the artificial and disembodied agents are experienced as "live" and capable to adapt to the user's politeness during runtime.

The continuous display of the AEA and PEA agents was interpreted differently by our participants. For some, it was an indicator that the system is online and ready. Others assumed that the device would be listening to commands non-stop. Consequently, they omitted the wake word and were irritated by being ignored from the attentive-looking agent. Moreover, users expressed discomfort of the agent starring at them during idle time. We noticed that the permanent presence of the agent leads to a feeling of constantly being observed by it. Some even felt like the agent would be waiting impatiently for the user until assistance is needed again.

We recommend hiding the agent between tasks to avoid social awkwardness and domestic intrusion. The reappearance can serve as feedback to indicate that the system recognized the wake word and is listening to commands. As another advantage of temporarily hiding the embodiment, other agents have the opportunity to re-embody the device, for example, for handling different task domains or assisting different users [34]. The transition could be implemented either as a fading effect or, for instance, with the agent walking in and out. In a playful context or in situations that require a continuous indicator of availability, such as for interactive public displays, the agent could alternatively be always visible but suggestive of being distracted. One participant proposed the agent being occupied reading a newspaper. As this might suggest to the user that the agent is busy and unavailable, we recommend trying more subtle deflections, for example letting the agent's gaze wander to the sides of the screen.

7 LIMITATIONS AND FUTURE WORK

The lower popularity of the prototype with the AEA is not necessarily due to the nature of digital rendering but might stem from our specific implementation. Users criticized the "creepy" appearance of the artificial agent and its lack of gestures. We assume that a higher technical sophistication could have improved its humanness. For this study, however, it was a deliberate design decision to compare different levels of realism. To exclude uncanniness effects, it would be insightful to replicate the presented experiment with a highly realistic rendered agent compared to a live-recorded human agent for distinguishing between effects from visualization and agency. Although the user experience might have benefited if the PEA had more human-like behaviour in terms of intonation, gestures, or facial expressions, we decided to match the artificial agent closely for higher comparability. For this study, we aimed for consistent agent realization, hence, we gave the AEA a computer-generated voice and the PEA a real human voice. Future work could investigate the impact of the AEA having a human voice and compare it to the PEA condition with a human voice.

We provided our participants with a predefined list of tasks. Some tested the system capabilities by asking additional questions. The VA responded that it cannot assist with that inquiry. Potentially, this could reduce the ecological validity. On the other hand, advantages of this established method for dialog system testing [45] are a structured procedure with high comparability and a feasible response preparation, as the questions are predictable. Further, the experiment covered mostly simple commands. More complex interactions, such as multi-step conversations, could also be investigated in future studies.

The experiment sample was skewed toward young ($M = 24.0$) men (78%) with a computer science background (42%). We cannot exclude an influence of this bias on the results concerning the agent's gender and age, or the participants' affinity towards technological innovations. Prior research suggested an effect of technical knowledge on the interactions with VAs [33]. Moreover, we decided to follow the conventional industry default of a female agent to avoid gender novelty as confounding factor. Since the focus of this research was on realism and the resulting artificiality of the agent, and not on gender comparisons, we did not further investigate this aspect with additional conditions. Therefore, we suggest future research to look at users' preferences for agents of different gender.

As the present study investigated short-term effects in a lab environment, it would be interesting to compare the results to a long-term exploration of different agent embodiments in a home setting – especially in terms of social presence, privacy concerns, and emotional bonding with the agent.

8 CONCLUSION

In this paper, we set out to understand the potentials and challenges of introducing agent embodiment to smart displays and compared different degrees of visual realism. Our contribution builds on a between-groups study with 60 participants. Using a Wizard of Oz method, we compared three conditions: no agent embodiment, artificial embodiment and photorealistic embodiment. In the quantitative system evaluation, we found similar user experience ratings across all conditions. Yet, the users had clear preferences and provided valuable insights on their views about the visualization and the permanent agent visibility. Moreover, we unexpectedly observed that the users were less polite towards the agent with photorealistic appearance. Our work identifies critical design considerations on how to embody voice assistant agents on smart displays to achieve a higher user satisfaction. The findings also provide orientation for researchers to quantitatively examine embodied smart display agents with targeted measurements.

ACKNOWLEDGMENTS

This work was partially funded by Klaus Tschira Foundation, by the FET-Open Project 951846 "MUHAI – Meaning and Understanding for Human-centric AI" funded by the EU program Horizon 2020, as well as the German Research Foundation DFG as part of Collaborative Research Center (Sonderforschungsbereich) 1320 "EASE – Everyday Activity Science and Engineering", University of Bremen (http://www.ease-crc.org/) conducted in subproject H02.

REFERENCES

[1] Elisabeth André. 2011. Design and evaluation of embodied conversational agents for educational and advisory software. In *Gaming and Simulations: Concepts, Methodologies, Tools and Applications*. IGI Global, Hershey, Pennsylvania, USA, 668–686.

[2] Sean Andrist, Michael Gleicher, and Bilge Mutlu. 2017. Looking Coordinated: Bidirectional Gaze Mechanisms for Collaborative Interaction with Virtual Characters. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*

(Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 2571–2582. https://doi.org/10.1145/3025453.3026033

[3] Wilma A Bainbridge, Justin Hart, Elizabeth S Kim, and Brian Scassellati. 2008. The effect of presence on human-robot interaction. In *RO-MAN 2008-The 17th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, IEEE Press, NJ, USA, 701–706.

[4] Erin Beneteau, Olivia K. Richards, Mingrui Zhang, Julie A. Kientz, Jason Yip, and Alexis Hiniker. 2019. Communication Breakdowns Between Families and Alexa. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, Article 243, 13 pages. https://doi.org/10.1145/3290605.3300473

[5] Timothy W. Bickmore and Rosalind W. Picard. 2005. Establishing and Maintaining Long-term Human-computer Relationships. *ACM Trans. Comput.-Hum. Interact.* 12, 2 (June 2005), 293–327. https://doi.org/10.1145/1067860.1067867

[6] Michael Bonfert, Maximilian Spliethöver, Roman Arzaroli, Marvin Lange, Martin Hanci, and Robert Porzel. 2018. If you ask nicely: A digital assistant rebuking impolite voice commands. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction : ICMI'18 : Boulder, CO, USA, October 16 - 20, 2018*. ACM, New York, NY, 95–102. https://doi.org/10.1145/3242969.3242995

[7] Sheryl Brahnam and Antonella De Angeli. 2012. Gender affordances of conversational agents. *Interacting with Computers* 24, 3 (2012), 139–153.

[8] J. Cassell, T. Bickmore, M. Billinghurst, L. Campbell, K. Chang, H. Vilhjálmsson, and H. Yan. 1999. Embodiment in Conversational Interfaces: Rea. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Pittsburgh, Pennsylvania, USA) *(CHI '99)*. Association for Computing Machinery, New York, NY, USA, 520–527. https://doi.org/10.1145/302979.303150

[9] Justine Cassell, Joseph Sullivan, Elizabeth Churchill, and Scott Prevost. 2000. *Embodied conversational agents*. MIT press, MA, USA.

[10] Susana Castillo, Philipp Hahn, Katharina Legde, and Douglas W. Cunningham. 2018. Personality Analysis of Embodied Conversational Agents. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents* (Sydney, NSW, Australia) *(IVA '18)*. Association for Computing Machinery, New York, NY, USA, 227–232. https://doi.org/10.1145/3267851.3267853

[11] Benjamin R. Cowan, Nadia Pantidi, David Coyle, Kellie Morrissey, Peter Clarke, Sara Al-Shehri, David Earley, and Natasha Bandeira. 2017. "What Can I Help You with?": Infrequent Users' Experiences of Intelligent Personal Assistants. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Vienna, Austria) *(MobileHCI '17)*. ACM, New York, NY, USA, Article 43, 12 pages. https://doi.org/10.1145/3098279.3098539

[12] Andreea Danielescu. 2020. Eschewing Gender Stereotypes in Voice Assistants to Promote Inclusion. In *Proceedings of the 2nd Conference on Conversational User Interfaces* (Bilbao, Spain) *(CUI '20)*. Association for Computing Machinery, New York, NY, USA, Article 46, 3 pages. https://doi.org/10.1145/3405755.3406151

[13] Munjal Desai, Kristen Stubbs, Aaron Steinfeld, and Holly Yanco. 2009. Creating trustworthy robots: Lessons and inspirations from automated systems. In *Proceedings of AISB Convention: New Frontiers in Human-Robot Interaction*. Society for the Study of Artificial Intelligence and the Simulation of Behaviour, Bath, UK, 49–56.

[14] Patrick Ehrenbrink, Seif Osman, and Sebastian Möller. 2017. Google Now is for the Extraverted, Cortana for the Introverted: Investigating the Influence of Personality on IPA Preference. In *Proceedings of the 29th Australian Conference on Computer-Human Interaction* (Brisbane, Queensland, Australia) *(OZCHI '17)*. ACM, New York, NY, USA, 257–265. https://doi.org/10.1145/3152771.3152799

[15] Jesse Fox, Sun Joo (Grace) Ahn, Joris H. Janssen, Leo Yeykelis, Kathryn Y. Segovia, and Jeremy N. Bailenson. 2015. Avatars Versus Agents: A Meta-Analysis Quantifying the Effect of Agency on Social Influence. *Human–Computer Interaction* 30, 5 (2015), 401–432. https://doi.org/10.1080/07370024.2014.921494

[16] Ted Grover, Kael Rowan, Jina Suh, Daniel McDuff, and Mary Czerwinski. 2020. Design and Evaluation of Intelligent Agent Prototypes for Assistance with Focus and Productivity at Work. In *Proceedings of the 25th International Conference on Intelligent User Interfaces* (Cagliari, Italy) *(IUI '20)*. Association for Computing Machinery, New York, NY, USA, 390–400. https://doi.org/10.1145/3377325.3377507

[17] Peter A Hancock, Deborah R Billings, Kristin E Schaefer, Jessie YC Chen, Ewart J De Visser, and Raja Parasuraman. 2011. A meta-analysis of factors affecting trust in human-robot interaction. *Human factors* 53, 5 (2011), 517–527.

[18] David Hanson, Andrew Olney, Steve Prilliman, Eric Mathews, Marge Zielke, Derek Hammons, Raul Fernandez, and Harry Stephanou. 2005. Upending the uncanny valley. In *AAAI*, Vol. 5. ACM, New York, NY, USA, 1728–1729.

[19] Marc Hassenzahl, Michael Burmester, and Franz Koller. 2003. AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität [AttracDiff: A questionnaire to measure perceived hedonic and pragmatic quality]. In *Mensch & Computer 2003*, Gerd Szwillus and Jürgen Ziegler (Eds.). B. G. Teubner, Stuttgart, 187–196.

[20] Álvaro Hernández-Trapote, Beatriz López-Mencía, David Díaz, Rubén Fernández-Pozo, and Javier Caminero. 2008. Embodied Conversational Agents for Voice-Biometric Interfaces. In *Proceedings of the 10th International Conference on Multimodal Interfaces* (Chania, Crete, Greece) *(ICMI '08)*. Association for Computing Machinery, New York, NY, USA, 305–312. https://doi.org/10.1145/1452392.1452454

[21] Guy Hoffman, Jodi Forlizzi, Shahar Ayal, Aaron Steinfeld, John Antanitis, Guy Hochman, Eric Hochendoner, and Justin Finkenaur. 2015. Robot Presence and Human Honesty: Experimental Evidence. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (Portland, Oregon, USA) *(HRI '15)*. Association for Computing Machinery, New York, NY, USA, 181–188. https://doi.org/10.1145/2696454.2696487

[22] Gilhwan Hwang, Jeewon Lee, Cindy Yoonjung Oh, and Joonhwan Lee. 2019. It Sounds Like A Woman: Exploring Gender Stereotypes in South Korean Voice Assistants. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. ACM, New York, NY, USA, Article LBW2413, 6 pages. https://doi.org/10.1145/3290607.3312915

[23] Gilhwan Hwang, Jeewon Lee, Cindy Yoonjung Oh, and Joonhwan Lee. 2019. It Sounds Like A Woman: Exploring Gender Stereotypes in South Korean Voice Assistants. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3290607.3312915

[24] Katherine Isbister and Patrick Doyle. 2002. Design and evaluation of embodied conversational agents: A proposed taxonomy. In *The first international joint conference on autonomous agents & multi-agent systems*. ACM, NY, USA.

[25] Rabia Khan and Antonella De Angeli. 2009. The attractiveness stereotype in the evaluation of embodied conversational agents. In *IFIP Conference on Human-Computer Interaction*. Springer, Springer, Heidelberg, Germany, 85–97.

[26] Sara Kiesler, Aaron Powers, Susan R Fussell, and Cristen Torrey. 2008. Anthropomorphic interactions with a robot and robot–like agent. *Social Cognition* 26, 2 (2008), 169–181.

[27] K. Kim, L. Boelling, S. Haesler, J. Bailenson, G. Bruder, and G. F. Welch. 2018. Does a Digital Assistant Need a Body? The Influence of Visual Embodiment and Social Behavior on the Perception of Intelligent Virtual Agents in AR. In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE Press, NJ, USA, 105–114. https://doi.org/10.1109/ISMAR.2018.00039

[28] Kangsoo Kim, Celso M de Melo, Nahal Norouzi, Gerd Bruder, and Gregory F Welch. 2020. Reducing Task Load with an Embodied Intelligent Virtual Assistant for Improved Performance in Collaborative Decision Making. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, IEEE Press, NJ, USA, 529–538.

[29] Bret Kinsella and Ava Mutchler. 2019. U.S. Smart Speaker Consumer Adoption Report 2019. https://voicebot.ai/smart-speaker-consumer-adoption-report-2019/

[30] Tomoko Koda and Pattie Maes. 1996. Agents with faces: The effect of personification. In *Proceedings 5th IEEE International Workshop on Robot and Human Communication. RO-MAN'96 TSUKUBA*. IEEE, IEEE, NJ, USA, 189–194.

[31] Michael Lankes, Regina Bernhaupt, and Manfred Tscheligi. 2007. An Experimental Setting to Measure Contextual Perception of Embodied Conversational Agents. In *Proceedings of the International Conference on Advances in Computer Entertainment Technology* (Salzburg, Austria) *(ACE '07)*. Association for Computing Machinery, New York, NY, USA, 56–59. https://doi.org/10.1145/1255047.1255058

[32] Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and evaluation of a user experience questionnaire. In *Symposium of the Austrian HCI and Usability Engineering Group*. Springer, Springer Berlin Heidelberg, Berlin, Heidelberg, 63–76.

[33] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf Between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. ACM, New York, NY, USA, 5286–5297. https://doi.org/10.1145/2858036.2858288

[34] Michal Luria, Samantha Reig, Xiang Zhi Tan, Aaron Steinfeld, Jodi Forlizzi, and John Zimmerman. 2019. Re-Embodiment and Co-Embodiment: Exploration of Social Presence for Robots and Conversational Agents. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (San Diego, CA, USA) *(DIS '19)*. Association for Computing Machinery, New York, NY, USA, 633–644. https://doi.org/10.1145/3322276.3322340

[35] Karl F MacDorman, Robert D Green, Chin-Chang Ho, and Clinton T Koch. 2009. Too real for comfort? Uncanny responses to computer generated faces. *Computers in human behavior* 25, 3 (2009), 695–710.

[36] Angie Lorena Marin Mejia, Doori Jo, and Sukhan Lee. 2013. Designing Robotic Avatars: Are User's Impression Affected by Avatar's Age?. In *Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction* (Tokyo, Japan) *(HRI '13)*. IEEE Press, NJ, USA, 195–196.

[37] Richard E Mayer and C Scott DaPra. 2012. An embodiment effect in computer-based learning with animated pedagogical agents. *Journal of Experimental Psychology: Applied* 18, 3 (2012), 239.

[38] Rachel McDonnell, Martin Breidt, and Heinrich H Bülthoff. 2012. Render me real? Investigating the effect of render style on the perception of animated virtual humans. *ACM Transactions on Graphics (TOG)* 31, 4 (2012), 1–11.

[39] Masahiro Mori et al. 1970. The uncanny valley. *Energy* 7, 4 (1970), 33–35.

[40] Masahiro Mori, Karl F MacDorman, and Norri Kageki. 2012. The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine* 19, 2 (2012), 98–100.

[41] Clifford Nass and Youngme Moon. 2000. Machines and mindlessness: Social responses to computers. *Journal of social issues* 56, 1 (2000), 81–103.

[42] Clifford Nass, Youngme Moon, and Nancy Green. 1997. Are machines gender neutral? Gender-stereotypic responses to computers with voices. *Journal of applied social psychology* 27, 10 (1997), 864–876.

[43] Clifford Nass, Jonathan Steuer, and Ellen R. Tauber. 1994. Computers Are Social Actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston Massachusetts USA) *(CHI '94)*. Association for Computing Machinery, New York, NY, USA, 72–78. https://doi.org/10.1145/191666.191703

[44] Catherine S Oh, Jeremy N Bailenson, and Gregory F Welch. 2018. A systematic review of social presence: Definition, antecedents, and implications. *Frontiers in Robotics and AI* 5 (2018), 114.

[45] Robert Porzel and Manja Baudis. 2004. The Tao of CHI: Towards Effective Human-Computer Interaction. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics: HLT-NAACL 2004*. Association for Computational Linguistics, Boston, Massachusetts, USA, 209–216. https://www.aclweb.org/anthology/N04-1027

[46] Alisha Pradhan, Kanika Mehta, and Leah Findlater. 2018. "Accessibility Came by Accident": Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. ACM, New York, NY, USA, Article 459, 13 pages. https://doi.org/10.1145/3173574.3174033

[47] Lingyun Qiu and Izak Benbasat. 2009. Evaluating anthropomorphic product recommendation agents: A social relationship perspective to designing information systems. *Journal of management information systems* 25, 4 (2009), 145–182.

[48] Byron Reeves and Clifford Nass. 1996. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge University Press, New York, NY, USA.

[49] Kristin E. Schaefer, Jessie Y. C. Chen, James L. Szalma, and P. A. Hancock. 2016. A Meta-Analysis of Factors Influencing the Development of Trust in Automation: Implications for Understanding Autonomy in Future Systems. *Human Factors* 58, 3 (2016), 377–400. https://doi.org/10.1177/0018720816634228 arXiv:https://doi.org/10.1177/0018720816634228 PMID: 27005902.

[50] Susanne Schmidt, Gerd Bruder, and Frank Steinicke. 2018. Effects of Embodiment on Generic and Content-Specific Intelligent Virtual Agents as Exhibition Guides.. In *ICAT-EGVE*. The Eurographics Association, Geneve, Switzerland, 13–20.

[51] Susanne Schmidt, Gerd Bruder, and Frank Steinicke. 2019. Effects of virtual agent and object representation on experiencing exhibited artifacts. *Computers & Graphics* 83 (2019), 1–10.

[52] Valentin Schwind, Katrin Wolf, and Niels Henze. 2018. Avoiding the Uncanny Valley in Virtual Character Design. *Interactions* 25, 5 (Aug. 2018), 45–49. https://doi.org/10.1145/3236673

[53] Jun'ichiro Seyama and Ruth S Nagayama. 2007. The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence: Teleoperators and virtual environments* 16, 4 (2007), 337–351.

[54] Selina Jeanne Sutton. 2020. Gender Ambiguous, Not Genderless: Designing Gender in Voice User Interfaces (VUIs) with Sensitivity. In *Proceedings of the 2nd Conference on Conversational User Interfaces* (Bilbao, Spain) *(CUI '20)*. Association for Computing Machinery, New York, NY, USA, Article 11, 8 pages. https://doi.org/10.1145/3405755.3406123

[55] Akikazu Takeuchi and Taketo Naito. 1995. Situated Facial Displays: Towards Social Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '95)*. ACM Press/Addison-Wesley Publishing Co., USA, 450–455. https://doi.org/10.1145/223904.223965

[56] Isaac Wang, Jesse Smith, and Jaime Ruiz. 2019. Exploring Virtual Agents for Augmented Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, Article 281, 12 pages. https://doi.org/10.1145/3290605.3300511

[57] Jun Xiao, John Stasko, and Richard Catrambone. 2004. An Empirical Study of the Effect of Agent Competence on User Performance and Perception. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems - Volume 1* (New York, New York) *(AAMAS '04)*. IEEE Computer Society, USA, 178–185.

[58] Beste F. Yuksel, Penny Collisson, and Mary Czerwinski. 2017. Brains or Beauty: How to Engender Trust in User-Agent Interactions. *ACM Trans. Internet Technol.* 17, 1, Article 2 (Jan. 2017), 20 pages. https://doi.org/10.1145/2998572

[59] Sean Zdenek. 2007. "Just roll your mouse over me": Designing virtual women for customer service on the web. *Technical Communication Quarterly* 16, 4 (2007), 397–430.

A APPENDIX: EXPERIMENT TASK LIST

The participants were provided with the following list as a printout and asked to take care of the tasks with the help of the voice assistant:

- Switch on the lights
- Turn on music (new "Fettes Brot" single)
- How long do eggs have to cook (they should be wax soft and are size M)
- Start a timer for the eggs
- Find out how many calories Gouda has
- Find out how many calories cashew cheese has
- Listen to science news
- Learn how the women's world cup final ended
- Find out if it's gonna rain today
- Check if there are appointments for today in the calendar
- Order a new micro USB charger (under 8€ and from Samsung)

The items were written out in German and English.

Publication P9

“Seeing the faces is so important”—Experiences from online team meetings on commercial virtual reality platforms

Michael Bonfert, Anke V. Reinschluessel, Susanne Putze, Yenchin Lai, Dmitry Alexandrovsky, Rainer Malaka, and Tanja Döring

This publication investigates VR team meetings in comparison to videoconferences. In a case study (N=32) during the COVID-19 pandemic, we conducted twelve lab meetings on five platforms over four months, generating authentic insights under genuine conditions. The results show that VR meetings better resemble in-person meetings regarding realistic group dynamics before and after the official meetings and the impression of being together in the same place. However, videoconferences are closer to reality regarding real faces and emotions, discreet side communication, and secondary tasks. They also increased the perceived involvement and co-presence. We discuss further results on spatial aspects, meeting atmosphere, expression of emotions, meeting productivity, and user needs.

My contribution: I had the leading role in contributing to the research idea (*conceptualization*) and study design (*methodology*). I made major contributions to the setup of the test environments (*software*), data acquisition (*data curation*), and qualitative data analysis (*formal analysis*). I created the majority of the figures (except the plots) and presentation material (*visualization*). I made a major contribution to the writing of the manuscript (*writing – draft, review, & editing*). I primarily coordinated the project (*project administration*).

Published in *Frontiers in Virtual Reality*, Section *Technologies for VR*, Research Topic *Everyday Virtual and Augmented Reality: Methods and Applications, Vol. II*, 2023

TYPE Original Research
PUBLISHED 11 January 2023
DOI 10.3389/frvir.2022.945791

OPEN ACCESS

EDITED BY
Daniel Zielasko,
University of Trier, Germany

REVIEWED BY
Mark Billinghurst,
University of South Australia, Australia
Myungho Lee,
Pusan National University, South Korea

*CORRESPONDENCE
Michael Bonfert,
bonfert@uni-bremen.de

SPECIALTY SECTION
This article was submitted to Technologies for VR,
a section of the journal
Frontiers in Virtual Reality

RECEIVED 16 May 2022
ACCEPTED 23 November 2022
PUBLISHED 11 January 2023

CITATION
Bonfert M, Reinschluessel AV, Putze S, Lai Y, Alexandrovsky D, Malaka R and Döring T (2023), Seeing the faces is so important—Experiences from online team meetings on commercial virtual reality platforms.
Front. Virtual Real. 3:945791.
doi: 10.3389/frvir.2022.945791

Seeing the faces is so important—Experiences from online team meetings on commercial virtual reality platforms

Michael Bonfert*, Anke V. Reinschluessel, Susanne Putze, Yenchin Lai, Dmitry Alexandrovsky, Rainer Malaka and Tanja Döring

Digital Media Lab, University of Bremen, Bremen, Germany

During the COVID-19 pandemic, online meetings became common for daily teamwork in the home office. To understand the opportunities and challenges of meeting in virtual reality (VR) compared to videoconferences, we conducted the weekly team meetings of our human-computer interaction research lab on five off-the-shelf online meeting platforms over 4 months. After each of the 12 meetings, we asked the participants ($N = 32$) to share their experiences, resulting in 200 completed online questionnaires. We evaluated the ratings of the overall meeting experience and conducted an exploratory factor analysis of the quantitative data to compare VR meetings and video calls in terms of meeting involvement and co-presence. In addition, a thematic analysis of the qualitative data revealed genuine insights covering five themes: spatial aspects, meeting atmosphere, expression of emotions, meeting productivity, and user needs. We reflect on our findings gained under authentic working conditions, derive lessons learned for running successful team meetings in VR supporting different kinds of meeting formats, and discuss the team's long-term platform choice.

KEYWORDS

CSCW, virtual reality, social VR, remote collaboration, virtual meetings, video conferencing, autoethnography, case study

1 Introduction

The COVID-19 pandemic required many people to work from the home office. Online meetings with video conferencing software became omnipresent after this trend had been promised for decades (Nilles, 1975). Although video calls provide many advantages, such as seeing meeting participants or allowing for screen sharing, they still yield limitations, such as restricted social interaction between participants. Social virtual reality (VR) platforms might provide beneficial alternatives for online team meetings by gathering everyone in one virtual room. Previous studies showed positive psychological effects of social VR platforms (Barreda-Ángeles and Hartmann, 2022) and that group behaviors and emotional responses to it are largely similar to face-to-face encounters (Moustafa and Steed, 2018). However, it seems that the generation of ideas as a creative process is hindered (Brucks and Levav, 2022). Beyond the faithful reproduction of in-person meetings, research further explores the possibilities of VR for enhancing social encounters and collaboration outside the restrictions of reality (Slater and Sanchez-Vives, 2016; McVeigh-Schultz and Isbister, 2021a).

With an increasing number of companies investing in immersive technologies and the future of the metaverse, social encounters in virtual environments (VE) will become commonplace in the foreseeable future. We can already observe this trend on platforms for

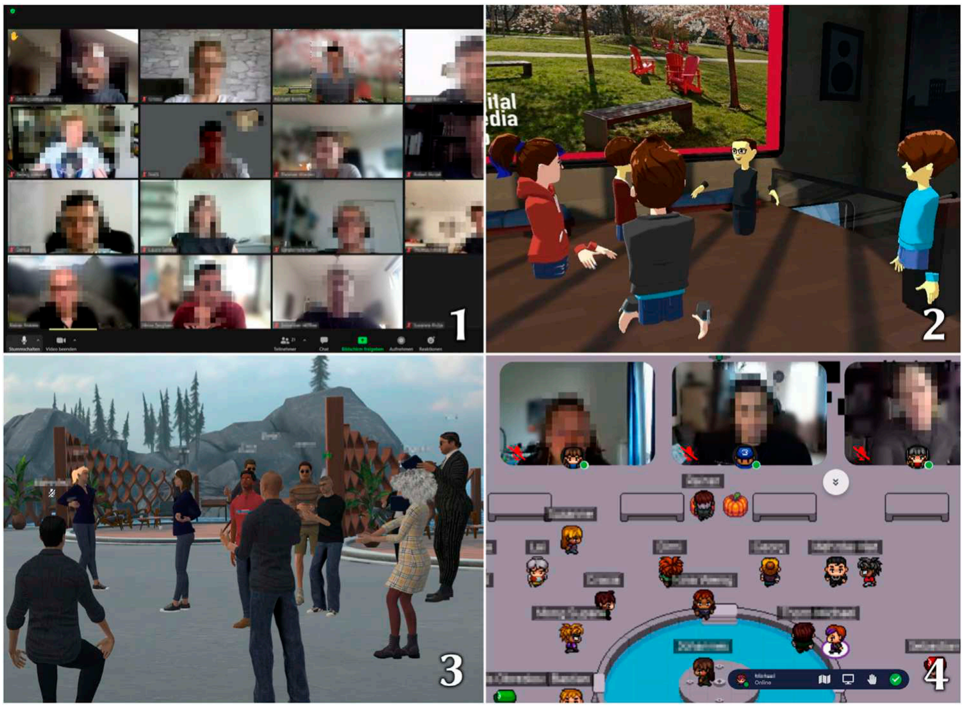


FIGURE 1
The off-the-shelf platforms compared in the wild for our weekly team meetings: (1) Our *status quo*, Zoom, compared to the two virtual reality platforms (2) AltspaceVR and (3) Engage, as well as the hybrid (4) Gather Town combining video feeds with a spatial environment. StarLeaf was also used in the case study but is not shown here as it has a similar interface to Zoom.

personal use where users meet in VR to socialize, play, and experience cultural events together (Sykownik et al., 2021)—even more pronounced during the pandemic (Rzeszewski and Evans, 2020). This development also extends to the business context. In 2022, 52% of employees were open to having meetings or team activities in the metaverse (Microsoft, 2022). The real estate company eXp Realty is operated entirely in virtual offices with more than 75,000 employees[1]. Numerous small companies and start-ups have launched collaborative platforms for meetings in virtual space. Also, large companies are preparing for a future in which professionals get together in virtual or blended realities, such as with Microsoft Mesh[2] or Horizon Workrooms by Meta[3]. Enabled by rapid technological development, immersive teamwork is more topical than ever. Still, many advertised features are visions of the future and blur the public perception of present possibilities.

To explore current opportunities and restrictions of commercial VR meeting software and to gain a deeper first-hand understanding of the advantages and drawbacks of authentic meetings in VR, we conducted the regular team meetings of our university's human-computer interaction (HCI) research lab in VR to compare the attendees' experiences with meetings on video conferencing platforms. This experimental shift from video-only platforms was intrinsically motivated, not by conducting this study. Therefore, the authors had the opportunity to independently evaluate the experiment, resulting in genuine insights from experiences in the wild. Over 4 months in 2020, we conducted twelve meetings on five different platforms–seven in VR and five on video conferencing or hybrid platforms for comparison. The different platforms are shown in Figure 1. We accompanied the team members from the expectations before the first VR meeting to the final conclusions on the long-term software choice by collecting feedback right after each meeting through online surveys. As part of the lab team, the authors also attended the meetings, which enabled an additional autoethnographic perspective to reflect on the meetings with different platforms.

With this, our case study adds to the growing body of research that investigates the gathering of people on various professional occasions on VR platforms. Previous studies explored attendees' experiences at professional social events in VR, e.g., at academic conferences and workshops (Erickson et al., 2011; Kirchner and Nordin Forsberg, 2021; Lahlou et al., 2021; Williamson et al., 2021), and group dynamics in social VR outside professional context (Moustafa and Steed, 2018; Scavarelli et al., 2021). However, in contrast to the existing body of work, this case study focuses on the participants' personal experiences during regular online team meetings in different mediums over an extended period. Thereby, it covers both VR and video conferencing solutions. Our study aims to explore whether currently available off-the-shelf social VR platforms meet the needs and preferences of the team members for attending weekly lab meetings as a competitive alternative to conventional video conferencing solutions. Based on 200 completed survey responses, we analyze ratings and statements of the meeting experiences and the different aspects that influenced these

1 https://www.virbela.com/customer-stories/exp-realty.

2 https://news.microsoft.com/innovation-stories/microsoft-mesh.

3 https://about.fb.com/news/2021/08/introducing-horizon-workrooms-remote-collaboration-reimagined.

experiences. Our analysis identifies central themes that matter for VR meetings and provides unique insights into the experiences of meeting on various platforms. While most of the individual themes have previously been discussed in the literature as isolated elements, encountering them integrated into authentic experiences of participants who compare it to videoconferencing in an in-the-wild study allowed us to assess and reconsider various factors, identify priorities, and add new aspects.

2 Related work

Research on computer-mediated collaboration has a long history. A broad landscape of literature provides an understanding of how people can work together remotely comprising various modalities, cultures, technologies, and goals of remote work (Galegher et al., 2013; Raghuram et al., 2019). Video-based meeting solutions received much attention in the late 1990s (Finn et al., 1997; Hinds, 1999) and provided strong indications that video-mediated meetings can have an equally good quality as face-to-face meetings (Olson et al., 1997). Still, for decades, videoconferencing was no competitive substitute for in-person meetings for large parts of the general public–or not explored as one–until an enforced shift due to the COVID-19 pandemic (OECD, 2021).

Along with the rise of videoconferencing platforms, the term "Zoom fatigue" was prominent in public and scientific discourse (Shockley et al., 2021). It describes the exhaustion after holding many or long meetings as videoconferences. Possible causes are suspected to be the cognitive load, always seeing oneself, and reduced mobility during video calls (Bailenson, 2021), but also spatial reduction of the conversation partners and their background to a 2D projection (Nadler, 2020) calling for mitigation or alternatives. According to Shockley et al. (2021), group belongingness was found to be the most consistent factor protecting from videoconference fatigue. Correspondingly, early research suggested that in remote collaboration, the employed technologies impact the communication outcome dependent on the interpersonal interactivity (Burgoon et al., 1999). While collaborating *via* videoconferences has been compared to other modalities, such as audio-only or in-person meetings (Ochsman and Chapanis, 1974; Daft and Lengel, 1986; Burgoon et al., 1999; Hinds, 1999; Bailenson, 2021), this work adds to the growing body of research with direct comparison to collaboration in VR. The following two sections will describe previous work on characteristics of social VR that might foster interpersonal interactivity and the feeling of group belongingness or mitigate limitations of videoconferencing in other ways. We further discuss how VR technology has previously been used for social encounters in a professional context.

2.1 Social interactions in VR

People use social VR platforms for a broad range of reasons. The strongest motives are meeting people, staying in contact, and experiencing social presence (Sykownik et al., 2021). Interacting in VEs offers considerable potential for high social presence (Short et al., 1976) as the spatial nature of the medium allows users to encounter other people in a shared space and affords complex social interactions and group dynamics. Social presence as "the connection of people *via* telecommunication systems" has been a central concept for comparing various forms of computer-mediated collaboration for decades (Nowak, 2001). Multiple factors affect how people interact with each other in VEs, including their visual and behavioral (self-) representation, the perceived agency of others' avatars, and potentially immersion (Kyrlitsias and Michael-Grigoriou, 2022). Different studies yield inconclusive results on the effect of visual realism on social presence (Nowak and Biocca, 2003; Bente et al., 2008; Kang and Watt, 2013; Zibrek and McDonnell, 2019) and imply dependence on other aspects of the simulation such as behavioral realism (Bailenson et al., 2005), which is generally a powerful predictor of social presence (Oh et al., 2018), such as turn-taking in a conversation (Bailenson et al., 2004). The social VR platforms tested in this case study differ considerably in the realism of virtual human representations introducing an interesting testbed with variance.

Further, sound is an important consideration for social interactions in VR. Spatial audio has been shown to impact the user's sense of presence (Poeschl et al., 2013; Kern and Ellermeier, 2020). Although it allows the user to identify the direction of different audio sources, concurrent talking of several people poses challenges for listeners in immersive environments. It can be effectively mitigated by helpful visual cues (Gonzalez-Franco et al., 2017). Besides the representation of virtual humans and verbal conversations, social VR interactions are influenced using non-verbal communication. On various social VR platforms, previous studies investigated the role of non-verbal cues and limitations in essential aspects of it and discussed similarities or discrepancies to non-verbal behavior patterns offline (Yee et al., 2007; Wigham and Chanier, 2013; Maloney et al., 2020; Tanenbaum et al., 2020). For instance, group arrangements, proxemics, and the preservation of personal space have been found to resemble offline behavior to a large extent (Bailenson et al., 2001; Yee et al., 2007; Hecht et al., 2019; Williamson et al., 2021) with adequate emotional responses to it (Wilcox et al., 2006). Also, previous research has shown that people can demonstrate similar social responses in virtual simulations as in reality. For example, in a study investigating compliance as done in the experiment by Milgram and Gudehus. (1978), the participants were equally or more compliant in VR compared to a control condition in physical reality (Dzardanova et al., 2021). Unfortunately, a problem that transfers from offline social interactions to virtuality, as well, is harassment (Freeman et al., 2022). As the communities are still establishing social norms for VR interactions, their governance remains a challenge (Blackwell et al., 2019a; Blackwell et al., 2019b; McVeigh-Schultz et al., 2019).

Altogether, many factors determine how a person in social VR perceives and interacts with another virtual human. For the collaboration of whole teams, however, not only the interaction between two people must be considered. Related work explored group dynamics and team communication in social environments that can resemble but also go beyond the possibilities of video conferencing. Torro et al. (2021) describe the impact of non-verbal communication and spatial information of social VR and why this makes it a game-changing medium for organizations. Due to the possibility of simulating any imaginable communication process, the authors argue that social VR has the potential to exceed the communication effectiveness of video conferencing and real-world settings. As one reason, they state how new forms of group dynamics can be facilitated and how teams benefit from formal and informal encounters in VR (Torro et al., 2021).

2.2 Meeting in VR

Researchers have explored the opportunities of remote meetings and social gatherings for a long time. Although the technical capabilities looked entirely different in 1999, the outcome of a study by Ståhl still shows similarities to current work. The most significant needs concerned the feedback on the users' connection in terms of visual attention, audibility, and network connectivity. In recent years, due to technological advancements and being incentivized by COVID-19, the body of literature on meetings and events in social VR is growing rapidly. Typically, the study designs in this research field are not lab-based and less controlled but of observatory nature and in the wild.

In this manner, Moustafa and Steed. (2018) conducted a longitudinal study on social interactions in small groups already known to each other moving their contact to social VR. The authors found that the participants experienced similar emotional states as in real-life socializing, which suggests high co-presence and transferability of existing group dynamics. Concerning educational purposes, a literature review explored interactions relevant in social xReality (XR) learning spaces and provided an overview (Scavarelli et al., 2021). For example, a study by Yoshimura and Borst. (2021) reports on class meetings experienced in VR comparing access with a head-mounted display (HMD) and in desktop mode. The students attended lectures and presented in VR. The experiences highly depended on how comfortable the attendees were with the HMD and if it made them sick. In another experiment, Ginkel et al. (2019) found a close resemblance between learning outcomes from training presentations in VR compared to face-to-face training.

Previous studies also explored the use of social VR environments in the academic community, such as social events at scientific conferences. While paper presentations can be effectively realized in videoconferences, it is much more challenging to enable virtual conference attendees to meet and connect with other participants. Yet, the informal exchange during coffee breaks and at receptions is essential to the success of academic collaborations. Therefore, organizers searched for virtual compensation, which was sometimes accompanied by scientific evaluation. For instance, Kirchner and Nordin Forsberg. (2021) organized a virtual conference where they held a reception on the VR platform Engage and performed a qualitative evaluation. The participants reported lively experiences but struggled with the spatial audio and discomfort from the HMDs. Further, the participants felt restricted in getting to know people they had never met before due to missing facial expressions. For fostering dynamic group conversations on similar occasions, Rogers et al. (2018) proposed displaying word clouds around groups hinting at the discussed topics to help strolling participants find a suitable group to join. Research by Williamson et al. (2021) analyzed the user proxemics during an academic workshop in VR. The results showed proxemic interactions between attendees that are congruent to physical settings and afford dynamic group formations dependent on properties of the VE. Beyond characteristics of personal space in social interactions known from reality, the possibility to enable participants to fly with their avatar added a dimension of user proxemics that was most notable during presentation situations.

A few studies investigated the social interactions at entire conferences in VR. The first evaluation of a virtual avatar-based conference by Erickson et al. (2011) comprised 500 attendees in Second Life[4] and was considered a "reasonably successful" event. According to the authors, the system must afford the formation of small groups for having focused interactions while breaking up and remixing into other groups over time. The analysis showed that the loud spatial audio disrupted conversation privacy and led to increased distances between groups which inhibited remixing. Structured social events worked better for the participants than more informal settings. In the end, none of the interviewed attendees experienced the virtual substitute and the face-to-face conference. A decade later, Lahlou et al. (2021) studied a conference in VR accompanied by video calls. The authors state two goals that a VR installation must meet for successful conferencing: (1) the development of knowledge and (2) informal social interaction. They find that current solutions still lack opportunities for natural social encounters and relational space. The researchers recommend careful preparation for organizers and suggest special consideration of onboarding processes and catering for socializing. In the same year, the conference IEEE VR 2020 was held entirely virtually for the first time. The accompanying evaluation by Ahn et al. (2021) provides a detailed comparison between the usage of different media platforms and their appropriateness for typical conference tasks. Again, the results point out the social constraints of the VE but also show advantages for connecting compared to other text-based platforms. Here, the social VR platform was rated as most appropriate for socializing and networking and providing the highest social presence. Most attendees who joined the VE decided to access it *via* desktop computer, although many owned an HMD. While the authors advocate for making use of the unique possibilities of virtual conferencing, they echo related research by advising caution not to transfer the substituted offline event directly to virtual space but to adapt the format and purpose of the event. In this spirit, McVeigh-Schultz and Isbister. (2021b) argue for conceiving solutions of VR collaboration that deliberately deviate from direct replication of familiar social encounters. Instead of imitating a face-to-face work environment, the authors promote using the full potential of immersive technologies to create an enriched experience of remote collaboration. The approach taken in this case study provides a flexible setting for this as the attendees could test how different platforms can serve the purposes of the meetings and reinvent the format along the way. Without predetermined system choices and autoethnographic insights, the team could adjust the conduct of the meeting every week over several months and compare the experiences.

During the COVID-19 pandemic, researchers were challenged with finding appropriate and practical scientific methods to continue conducting studies (Nind et al., 2021). Research on computer-mediated communication has established various methods and tools for evaluating specific aspects of the behaviors and opinions of interlocutors, e.g., turn-taking and behavioral analyses to examine meeting dynamics (Samrose et al., 2021; Samrose et al., 2018) and simulated conversations (Abdullah et al., 2021), audio-visual capturing to understand communication characteristics (Byun et al., 2011; Koseki et al., 2020), linguistic analyses (Kramer et al., 2006; Fägersten, 2010), self-reports in interviews (Bleakley et al., 2022) and standardized questionnaires (Nowak and Biocca, 2003), autoethnographic methods (Mack et al., 2021), or a self-hosted and modified version of a social VR platform to understand the proxemics (Williamson et al., 2021). For this research, we rely on the conjunction of quantitative and qualitative methods from self-reports in a survey that includes standardized questionnaire items and open questions. Our approach allows longitudinal analysis through unique identifiers while preserving anonymity. Previous literature investigated the use of

4 https://secondlife.com.

TABLE 1 Overview over the 12 evaluated meetings.

ID	Medium	Participants	Responses	HMD Users	Duration (min)	Presentations
M1	Starleaf	20	18	n/a	41	No
M2	Altspace	19	16	87.5%	59	Yes, one
M3	Altspace	22	19	84.2%	53	Yes, one
M4	Altspace	20	19	84.2%	21	No
M5	Altspace	22	18	72.2%	21	No
M6	Altspace	20	16	68.8%	42	Yes, one
M7	Engage	18	15	93.3%	57	Yes, three
M8	Engage	15	11	72.7%	18	No
M9	Zoom	20	17	n/a	43	No
M10	Zoom	24	18	n/a	59	Yes, one
M11	Gather Town	22	18	n/a	84	Yes, one
M12	Zoom	17	15	n/a	83	No

FIGURE 2
This shows the setup of the *AltspaceVR* platform. It was an indoor room with large windows and a presentation screen.

VR technology for conducting conferences or socializing events, focusing on the comparison to the face-to-face equivalent. On the other hand, this work explores the potential of holding team meetings in VR and compares it to video conferencing while investigating, in both cases, the similarity to face-to-face meetings. Instead of determining usability issues of individual social VR platforms, as previous research has done systematically (Liu and Steed, 2021), the personal experiences and needs of the meeting participants are of central interest in this case study.

3 Methods

Over 4 months, from August until December 2020, we evaluated twelve weekly meetings on different platforms (see Table 1 for an overview) by inviting the participants to fill out a questionnaire after each session. We used a mixed-method approach and collected quantitative and qualitative data to investigate the participants' ratings of each meeting experience and their personal impressions.

3.1 Participants

Generally, the weekly meeting is attended by all team members, including technical and administrative staff, undergraduate and Ph.D. students, postdocs, professors, and guests. This leads to a heterogeneous sample concerning technical literacy, previous experience with VR, and expectations towards the meetings. Participation was voluntary and strictly anonymous, so we did not collect any demographic data. Nevertheless, the participants rated their prior experience using three items: How often they use VR in general, how often they use multi-user VR applications, and whether they use VR as part of their work. From these three items, we calculated an overall prior experience score (min 0 to 10 max). On average, the 18 participants of meeting M1 had a prior experience score of 3.167 (SD = 2.41). All group members had access to VR hardware, i.e., HTC Vive (Pro), Valve Index, or Oculus Quest 1. In total, we had 239 meeting participations in the 12 meetings with 32 distinct attendees completing at least one questionnaire. Four participants attended and filled out the questionnaires for all 12 meetings. 12 participants responded after at least 10 meetings. In each meeting were between 15 and 24 attendees. We did not collect demographic data such as age or gender because this would disclose the participants' identities due to the small group size. For the same reason, we did not ask about the HMD model used.

3.2 Evaluated meetings

The main purpose of the weekly team meetings was to report on and discuss current matters concerning the lab or its associates, which the group manager and administrative staff mostly did. Additionally, there

FIGURE 3
This shows the setup of the *Engage* platform. It was an outdoor scenario similar to an amphitheater with a presentation screen.

were presentations by undergraduates and group members (in 50% of the reported meetings, at least once on each platform). At the end of the meetings, individuals and small groups would discuss matters with the group manager irrelevant to the whole team, while everybody else would already leave. On average, the meetings investigated in this study lasted 48.42 (SD = 22.02) minutes and had 19.91 (SD = 2.47) attendees with details shown in Table 1. Before the COVID-19 pandemic, the lab meetings were held in person. Between March 2020 and the beginning of this case study, they were on Zoom or StarLeaf.

3.3 Procedure and questionnaire

The meeting was joined by the attendees' device of choice–an HMD (if applicable) or desktop client–and followed its regular structure. During the weekly meetings, the group discussed which platform to try next and when to change it. After each meeting, the researchers emailed all group members and guests linking to a Google Forms questionnaire. Each recipient, including the authors, could individually decide to take part in the survey or not, which resulted in an average response rate of 83.7% (SD = 6%). We made every effort to ensure there was no social pressure to participate and no fear of negative consequences for answering honestly. The questionnaire started with a consent form and a unique identifier (ID) chosen by each participant to identify repeated participation and allow longitudinal analysis anonymously. We asked how the meeting was accessed and found that HMDs were used in 92 out of 114 VR platform cases (80.7%). For the first session, we also assessed the prior experience with VR and expectations for the VR sessions. The questionnaire incorporated questions from the Social Presence Questionnaire by Nowak and Biocca. (2003), the User Burden Scale (Suh et al., 2016) and self-designed questions to answer on a 7-point Likert scale. Additionally, we asked open questions about the meeting experience, e.g., group interactions, comparisons to the other platforms, and memorable experiences. The questionnaire after the first and last session had a few additional items. After the first meeting, we asked about the participants' general usage of VR, their usage of multi-user VR applications, whether they work with VR, and their expectations

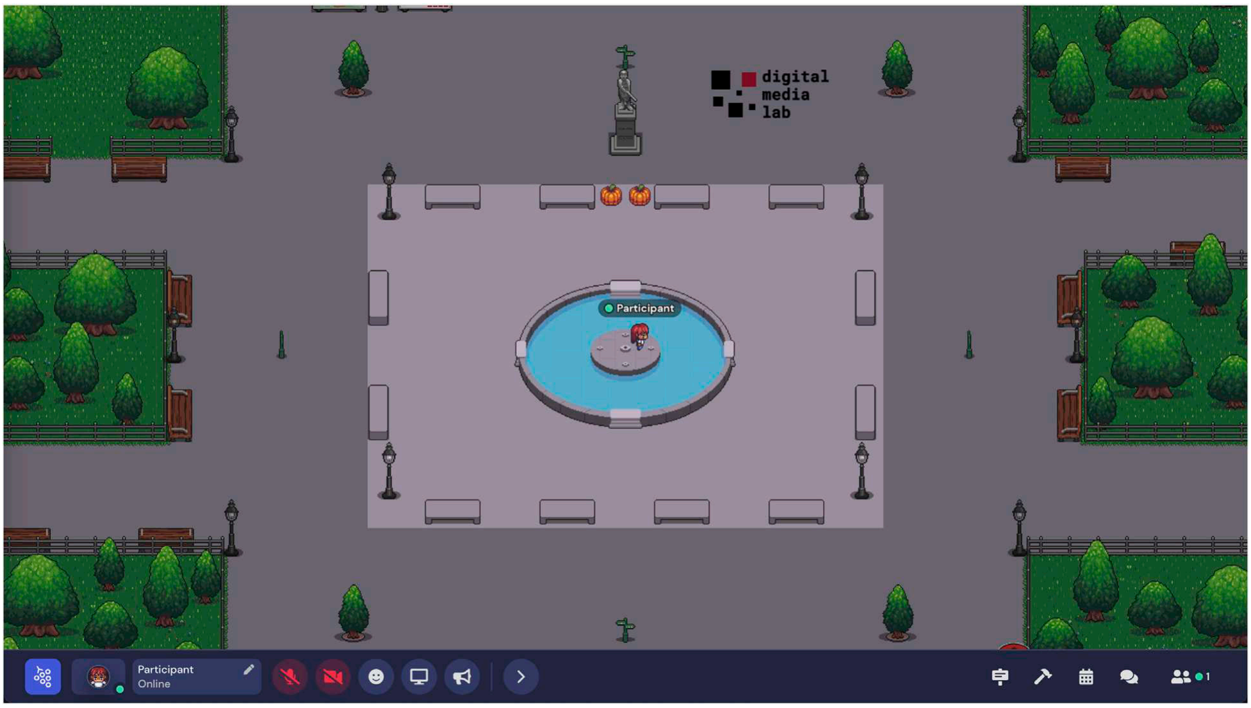


FIGURE 4
This shows the setup of the *Gather Town* platform. The scenery was an outdoor place in a park without a dedicated presentation screen. The lighter grey space identifies one coherent meeting area.

and worries for VR meetings. Additionally, the final survey after the last session incorporated questions about the preferred meeting platform and the reasons for the preference. In the supplemental materials, we provide an overview of all questionnaire items and, if applicable, their respective origins from standardized instruments.

3.4 Platforms

To gain insights into the suitability of openly available social VR services for team meetings in the wild, we solely used established and sophisticated platforms that allow integrating presentations and giving talks. The most critical factor for the software selection was the accessibility from all devices used by the attendees, as they were free to choose which device they preferred to access each meeting. Throughout the case study, we tested the VR platforms *AltspaceVR*[5] (see Figure 2, meetings 2–6) and *Engage*[6] (see Figure 3, meetings 7 and 8), the videoconference platforms *Zoom*[7] (see Figure 1 (1), meetings 9, 10 and 12) and *StarLeaf*[8] (meeting 1), as well as *Gather Town*[9] (see Figure 4, meeting 11), a hybrid of a virtual 2D environment and video calls. We provide additional screenshots from the meetings and a video figure in the supplemental materials.

The virtual setup of AltspaceVR was based on a template provided by AltspaceVR. We used a template from the category "Talk Show" because it had a big screen for presentations and consisted of a room with large windows. This layout was inspired by the real setting the meetings were held in pre-Covid. We modified the virtual space with the group's logo as visible in Figure 2. The table also resembled the U-shaped setup of our physical meeting room and gave the virtual room some structure. Participants could customize their avatar representation, mute themselves, react with emoticons, and move around the space. There was a room admin who could enable the megaphone feature for single persons to make them audible to the group in the whole room.

Similar to the setup of AltspaceVR, the meeting space of Engage was based on a template provided by the platform. The room design resembled an ancient outdoor forum with a U-shaped area to sit down to watch a stage. The stage featured a large presentation screen in the background. The meeting area also featured some educational exhibits and was placed on an island, viewing the Golden Gate Bridge in the distance. In contrast to the AltspaceVR settings, the people could actually sit down and were evenly spaced in the VE, creating the impression of a seated audience for the person on stage.

We relied on a pre-existing template for the Gather Town meeting space again. We chose a park setting (see Figure 4) featuring areas for socializing with games. We modified the template so that the area around the fountain was one meeting space–which is colored in light grey in Figure 4, where everyone inside could see the video of everyone else in this area (*cf.* Figure 1 (4)). Additionally, we added the two pumpkins between the upper benches to mark two tiles with a broadcast feature. The participants standing on these tiles were audible in the whole environment, similar to the megaphone feature of AltspaceVR. Outside of dedicated meeting areas, the visibility and audibility depended on the proximity of the participants, allowing the creation of small spontaneous groups.

3.5 Data analysis

The questionnaires collected quantitative data (7-point Likert scales) and qualitative (answers to open questions). The quantitative data were analyzed using factor analysis. As the questionnaires included various attributes that are not independent of each other, an exploratory factor analysis (EFA) with Varimax rotation on all questionnaire items was used to reveal hidden *factors* that group multiple related questionnaire items. This data-driven process allowed us to identify groups of questions that influenced the users' ratings of their meeting experiences. The qualitative data of the questionnaires were analyzed using thematic analysis by identifying similarities and inductively creating themes. A team of four researchers individually coded three responses, and after that, they discussed and iterated the resulting codes until they agreed on a common code system. This system was used to code all qualitative responses once. We received 200 responses. 155 of them contained qualitative data in addition to the questionnaire items. After the additions to the code system from the first cycle were discussed, a different researcher coded all responses a second time. This process resulted in 1,228 codings using 107 codes.

4 Results and lessons learned

In the following, we provide our analysis of the quantitative data collected in subsection 4.1 as well as insights on the five themes derived from the qualitative data in subsections 4.2 to 4.6. The questionnaire responses to the Likert scale questions are evaluated with a focus on the ratings of the meeting experience. To provide authentic impressions and derive and discuss dominating themes that explain and contextualize the quantitative results, we present the qualitative insights in greater depth in the following. We report on the experiences and statements of the participants on the five main themes that the thematic analysis revealed: Spatial Aspects, Meeting Atmosphere, Expression of Emotions, Meeting Productivity, and User Needs. Each subsection concludes with a Lessons Learned paragraph discussing central insights and linking them to related work.

4.1 Ratings of the meeting experience

As part of our post-meeting questionnaire, we obtained feedback on 19 Likert-scaled question items, shown in Table 2. In the following, we present the results of our quantitative data analysis of the users' ratings on those items.

5 https://altvr.com.
6 https://engagevr.io.
7 https://zoom.us.
8 https://starleaf.com.
9 https://gather.town.

TABLE 2 Overview of the most important questionnaire items. Questions from standardized instruments are denoted with *SPQ* for the Social Presence Questionnaire by Nowak and Biocca. (2003) or *UBS* for the User Burden Scale by Suh et al. (2016). The constellations of all five questionnaires is detailed in the supplemental material.

Attribute	Question
face_to_face_meeting SPQ	To what extent was this like a face-to-face meeting?
same_room_with_partner SPQ	To what extent was this like you were in the same room with your partner?
partner_realism SPQ	To what extent did your partners seem "real"?
choose_to_persuade SPQ	How likely is it that you would choose to use this system of interaction for a meeting in which you wanted to persuade others of something?
get_to_know_extent SPQ	To what extent did you feel you could get to know someone that you met only through this system?
meeting_involvement	I felt involved in today's meeting
active_participation	I participated actively in today's meeting
felt_noticed	I felt noticed by the other participants
felt_group_membership	I felt like being part of the group
meeting_productivity	We had a productive meeting today
worry_information_leak UBS	< The meeting platform > accesses and uses the device's microphone and camera. I'm worried about what information is being passed on by it
privacy_trustworthy UBS	< The meeting platform > 's policy about privacy is not trustworthy
privacy_requirement UBS	< The meeting platform > requires me to do a lot to maintain my privacy within it
moderation_quality	The moderation and guidance by the moderator in today's meeting were
agenda_structure	The agenda and structure of today's meeting were
vr_meetings_replace	I think the VR meetings will replace the video conferences in our lab
technical_difficulties	We had many technical difficulties
usage_confidence	I felt confident about using the system
total_meeting_experience	Overall, how was your meeting experience today?

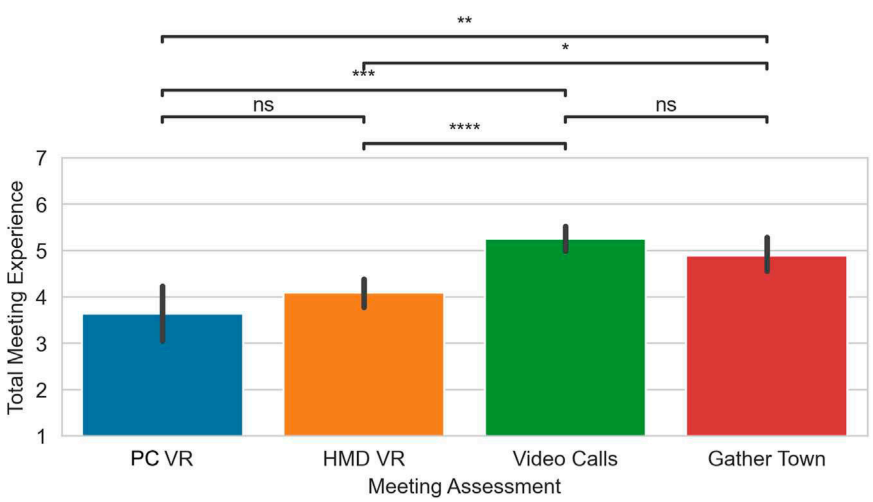


FIGURE 5
Barplot of the total meeting experience assessed in meetings M2–M12 together with the results of Bonferroni-corrected paired t-tests.

Starting from meeting M2, we asked the attendees about their overall meeting experiences on a 7-point Likert scale (poor 1 to 7 great). Figure 5 shows a barplot with the results of the total meeting experience. With a one-way ANOVA, we found a significant effect of the meeting modality on the total meeting experience ($F(3, 178) = 11.041, p < 0.01, \eta_p^2 = 0.157$). Bonferroni-corrected *post*

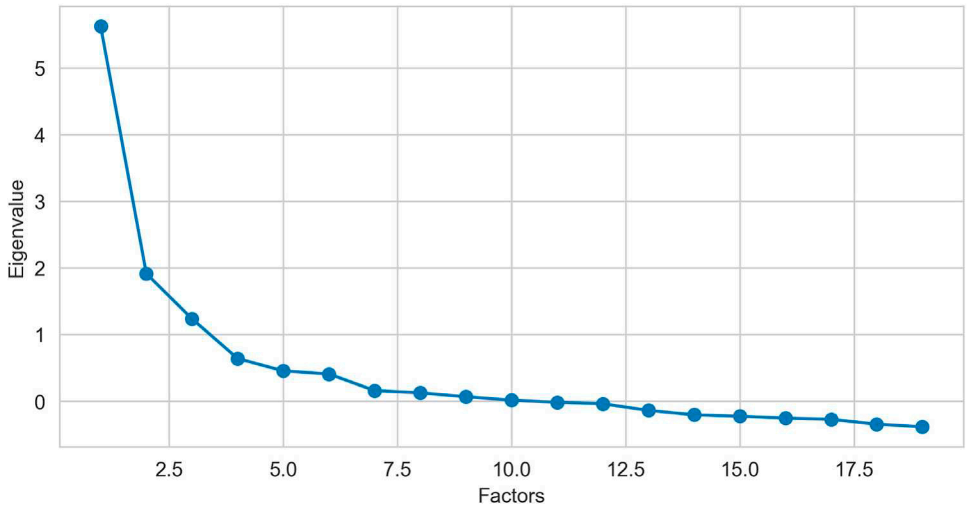


FIGURE 6
Scree plot of variance explained by each factor, resulting from the factor analysis on questionnaire items.

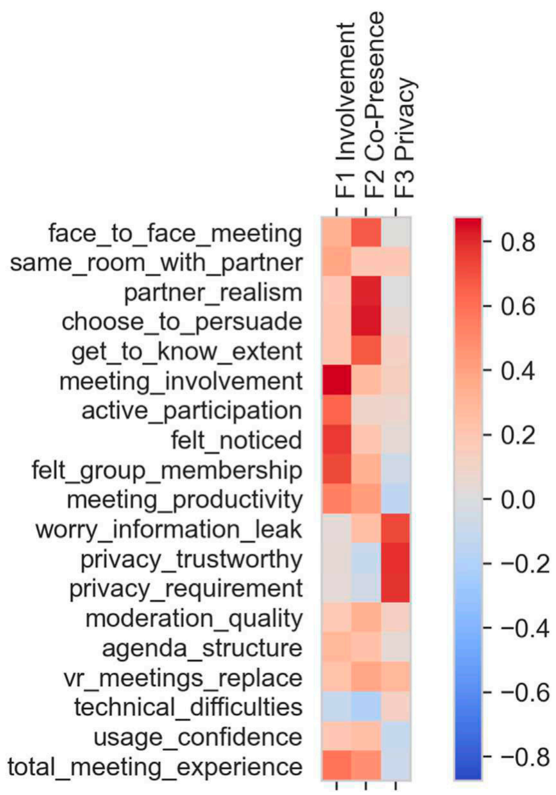


FIGURE 7
Loadings of all attributes (larger absolute values represented in darker colors) for the three factors (F1–F3) in the exploratory factor analysis.

hoc tests confirmed that the total meeting experience in the video call meetings was significantly higher than in the VR meetings for both VR options, VR *via* desktop PC (t (30.454) = −4.876, p < .01, d = −1.429) and VR using an HMD t (137.635) = −5.340, p < .01, d = −.821). For Gather Town, we also observed an improved meeting experience for PC VR (t (35.009) = −3.509, p < .01, d = −1.062), and HMD VR (t (45.499) = −3.124, p = .019, d = −.537), and no difference for the video call platforms (t (35.186) = −1.461, p = .917, d = −.372). These results indicate an advantage of video calls and the hybrid platform Gather Town compared to the VR conditions. Nevertheless, we did not observe any difference between the VR options t (35.156) = −1.321, p = 1.00, d = −.290) and, therefore, we did not differentiate between these two VR variants in the following quantitative analysis.

Since the attendees had a very diverse pre-experience with VR technology, we investigated whether the pre-experience, as assessed in meeting M1, correlates with the meeting experience in the first VR meeting M2. Using a Spearman rank correlation, we found that the two variables were moderately correlated (Spearman's r (13) = .514, p = .072).

However, the total meeting experience rating did not provide a conclusive impression of the users' perception of the meetings. Only four participants attended all twelve meetings, which prevented a more detailed within-subject analysis in the following. Still, we wanted to investigate exploratively which attributes contribute to the user ratings to understand the observed differences.

4.1.1 Exploratory factor analysis

To explore the structure of user ratings and identify possible factors in the multiple questionnaire items of Table 2, we conducted an exploratory factor analysis on the additional feedback obtained through the Likert scale questionnaire items. Figure 6 presents a scree plot of the variance associated with each factor, representing the explained variance for each resulting factor. This helps us to obtain the number of factors. Factors with high levels of explained variance are important, interpretable contributors to the overall model, whereas later factors explain less variance. The scree analysis revealed three factors with an Eigenvalue above 1. Continuing with these three factors, we calculated the factor loadings. Figure 7 shows the loadings of all question items for the three factors. The loading of an attribute for a factor can be interpreted as the correlation between the question item and the factor. Therefore, the higher the absolute loading value, the stronger that attribute is tied to that factor and predicted by it. On the other

FIGURE 8
A situation in Engage before the meeting where the participants grouped (left side) and one person walked up to this group (right side, black shirt).

hand, a loading of close to 0 indicates the absence of a (linear) relationship between the attribute and the factor.

This allowed us to find interpretations of these factors. Factor F1 loads highest on the questions related to *Involvement*, F2 corresponds to *Co-Presence* items, and F3 to items around *Privacy*. Subsequently, we compare VR meetings, video calls, and Gather Town with each other according to the three factors. With a one-way ANOVA, we found a significant effect of the meeting modality on factor *F1 Involvement* ($F(2, 156) = 4.263$, $p = 0.016, \eta_p^2 = 0.052$). Bonferroni-corrected *post hoc* tests confirmed a difference between VR and video calls with $t(126.17) = -3.195$, $p < .01$, $d = -.515$ with higher ratings for video calls. Another one-way ANOVA revealed a significant effect of the meeting modality on factor *F2 Co-Presence* as well ($F(2, 156) = 26.934, p < 0.01, \eta_p^2 = 0.257$). Post-hoc tests confirmed a difference between VR and video calls with $t(137.940) = -8.228$, $p < .01$, $d = -1.269$, and between Gather Town and video calls with $t(16.149) = -3.775, p < .01, d = -1.446$ each with better ratings for video calls. We could not find an effect for *F3 Privacy* ($F(2, 156) = 0.309$, $p = 0.735, \eta_p^2 = 0.004$), indicating conclusive experiences on that factor independent of the meeting modality. These results demonstrate the advantages of video calls in contrast to the other platforms.

4.1.2 Lessons learned

Concerning quantitative data, we could observe that video meetings outperformed VR meetings in terms of the total meeting experience. Furthermore, for two of the three factors revealed with the exploratory factor analysis, i.e., *F1 Involvement* and *F2 Co-presence*, we also found that video conferencing was ranked higher than VR meetings. Users tended to feel more involved and have higher co-presence with their colleagues in video calls than in VR meetings. With the immersive nature of VR and its capability to substitute the user's physical environment, it is surprising that the attendees had a stronger impression of others being present on video calls than in VR. To obtain a deeper understanding of the reasons for this assessment, we provide detailed insights on the experiences along the five themes that evolved from our qualitative analysis in the following subsections.

4.2 Spatial aspects

An essential difference between video calls and virtual 3D environments is the spatial component that allows movement in space and varying proximity to others. It influences the group dynamics and the audio considerations.

4.2.1 Movement

Before the study, P08 expected "more fun and more movement during the meetings" in VR. Indeed, the possibility of moving around in the VE led to more vivid and "a lot more dynamic" (P18) meetings compared to both video calls and physical meetings. Some attendees preferred that everyone be seated for the meeting to be more "orderly" (P28). Other participants appreciated "being able to move around and arrange in the room" (P12) and to "walk in front of the screen and use [their] body" (P18) during presentations. This was sometimes perceived as troublesome during presentations because avatars of other attendees were blocking the view onto the slides or the speaker: "All the time there was someone in my way" (P09). However, for presenting, the participants described advantages: "I also liked the more true-to-life presentation environment with the projector screen and with the ability to see the audience spread out in a semicircle instead of a gallery view like on Zoom" (P07). Many attendees highlighted the value of meaningful movements such as waving, gaze direction, and pointing as indicators, e.g., to initiate conversations or for turn-taking, making it "more personal" (P16). That "you could walk up to someone to talk to them" (P17) was perceived as more effortless and more natural (see Figure 8). Another use of the spatial component was to form queues, especially after the meeting when several attendees wanted to speak to the same person.

4.2.2 Proxemics

Similarly to movements, proxemics was used with purpose or meaning. "The physical arrangement also plays a role–who sits next to whom", mattered for P12, or when everyone in a circle is supposed to say something in turn. At the same time, the spatial relations did not always translate well to the VE: "In physical meetings, it is nice that you actually sit next to others. In VR, it still feels a bit odd (am I too close? Too far away?)—it is also a bit more difficult to exactly place yourself the way you would like" (P12). Other attendees also reported that they "felt awkward" (P05) or "felt bothered because the others came too close" (P16).

4.2.3 Group dynamics

The added spatial component of VEs allowed to form and vary group constellations and sizes, which was a crucial benefit to a majority of the participants: "The VR environment makes it much easier to converse in small groups and easily switch between talking to different people", explained P07. P06 pointed out that especially "before and after the meeting [interactions] were a lot easier and closer to in-person interactions than they would be in a Zoom meeting." The participants described a large variety of interaction types, including 1-to-all as in presentations, 1-on-1, small groups, large groups, the whole group, and dynamic transitions between the constellations. Other than in Zoom, small groups could split up and follow up on meeting topics in parallel, which is "barely possible [. . .] next to the official round" (P12) without blocking the meeting room for all others in a video call. Therefore, VR meetings were described as more dynamic, allowing more attendees to say something and contribute. On the other hand, Zoom meetings were characterized as "static" (P14).

VEs also allow the users to position themselves in a third dimension: elevation. This affords extended group arrangements and was primarily used for presentations and other 1-to-all communication. One participant experienced the group manager hovering mid-air above the rest of the team as a status imbalance: "It was beneficial that we could all see him. But it also gave the impression that he was 'superior' and talking down on us" (P18).

4.2.4 Spatial audio

The systems based on a VE are designed so that a conversation is only audible to users standing close by. The further away a listener is, the quieter the sound becomes. While this enabled several small groups to have private conversations simultaneously in the same room, it was perceived as cumbersome for the plenum when everybody should hear everybody else by default: "In case of a presentation or group discussion, sometimes you do not understand others and we used the 'megaphone' [to amplify voices in AltspaceVR] all the time" (P05). Furthermore, attendees reported that, also in VR, "the main questions you hear are: 'am I muted?' or 'can you hear me?'. The spatial audio is completely useless." (P05). Manually equipping speakers with the 'megaphone' as an amplifier was referred to as "a burden" (P01), not only by the participants who facilitated the meeting but also by observers. The thresholds for the spatial audio were often experienced as inadequate resulting in unnatural behavior: "Especially because of the spatial audio, people gathered closely in clusters so they can hear each other better. In reality, it would not be that condensed, and also not so hard to hear each other" (P18). Instead, P05 suggested "to have regular audio [for the plenum] and for private conversations sound bubbles."

4.2.5 Lessons learned

The VEs allowed the attendees to gather in a virtual workspace. Videoconference platforms, such as Zoom or Skype, recently introduced features to create a similar impression by visually cropping attendees and stitching them together in a shared room[10]. However, this lacks intradiegetic possibilities to utilize this added spatial component, for example, the eye gaze in turn-taking. When designing a VR meeting platform, we consider it essential to foster dynamic interpersonal interactions by supporting meaningful movements, such as approaching others, forming groups and queues, waving, pointing, or nodding. Non-verbal cues strongly impact social interactions in VR but have not yet tapped the rich potential familiar from offline behavior (Maloney et al., 2020; Tanenbaum et al., 2020). In line with related research, our participants reported spatial group arrangements and proxemics similar to in-person meetings. As Williamson et al. (2021) point out, elevation provides additional opportunities for an audience in VEs to arrange for better visibility. While this can be beneficial, we advise caution when activating flying features, as they might diminish professionalism or affect the perception of social status differences.

For flexible and realistic group dynamics, purposeful and configurable spatial audio with carefully chosen volume reduction parameters is required, as well as an option to hear everyone in a plenum. We observed similar behavior of our participants as in a study by Williamson et al. (2022), where users seemed to move close together to hear each other best despite loud background conversations. As our participants proposed, allowing users to create a private sound bubble with visible boundaries to shield their group might be helpful. Already in 1999, (Ståhl, 1999), reported in a similar case study that the uncertainty of being audible was one of two critical challenges for VR meetings. Even today, we still observe the need for more explicit indicators of one's comprehensibility within the room.

4.3 Meeting atmosphere

We further observed that the spatial component strongly influenced the meeting atmosphere. In the questionnaire before the first meeting in VR, P18 shared that the "feeling of working in the same place is missing" during the long home office directive. The potential of immersively feeling "like being somewhere else" (P12) for the meetings promised relief to some participants.

4.3.1 Co-presence and interpersonal interactions

Numerous attendees expressed excitement about "being in a room together" (P06) and "chatting 'face-to-face'" (P12). For P18, in the VE, it felt more like "coming together or gathering with the whole team" in a shared virtual office space than in video calls. P23 and P09 were reminded of physical meetings, which included that "interactions felt more like a real meeting" (P09). The effect was perceived to be stronger when accessing the VE with an HMD: "I felt more immersed and part of the group" (P09). P25 appreciated how "one has a better awareness for the group, such as closeness or distance, attention or distraction, *etc.*, which is missing on Zoom." While in VR, participants appreciated seeing everyone around them at a glance, the videoconference tools were repeatedly criticized for not displaying all attendees in large meetings: "I frequently do not register that they are attending the meeting" (P12). This is especially problematic when presenting as it "gives a wrong impression about the number of participants" (P11).

P18 appreciated being "surrounded by my colleagues, but without faces, it was not so personal. Like a coat of secrecy around everyone." With the cartoony avatars on AltspaceVR, to P23, it "felt like speaking with very intelligent and responsive game characters". This dilemma

10 https://blog.zoom.us/introducing-zoom-immersive-view https://www.skype.com/en/blogs/2021-05-together-mode.

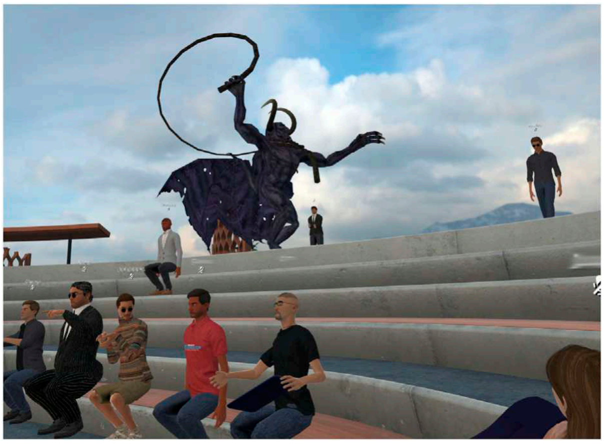


FIGURE 9
A user inserted a daemon to the scene whipping a colleague's avatar in the middle of a student's presentation.

was especially noticeable for guests. Several team members expressed the wish to see a guest's face "to get an idea of how [they are] as a real person" (P28). On the other hand, the avatar representation allowed guests to *"visit"* the meeting as equal to the other members.

4.3.2 Professional or playful

The initial atmosphere of the meeting was established mainly by the visual appearance of the interface, the avatars, and the environment. The setting of AltspaceVR was perceived as "sometimes too cartoonish to be taken seriously [. . .] more like in-game conversations" (P23). On the other hand, Engage was perceived as more formal and "much more orderly than in Altspace because we were seated" (P28). Both the professional and the playful styles have received positive and negative feedback from the participants. Some participants liked the professional atmosphere that Zoom or Engage created because it is similar to physical meetings and makes it easier to talk about serious topics. But it was also perceived as more boring and formal, and therefore some rejected the approach of Engage: "I really hated Engage. The uncanny avatars, technical difficulties, and overall 'seriousness' of the application did not work for me" (P09). Due to the less formal setting in AltspaceVR, "even the student was completely chilled–usually on Starleaf they're more nervous and quiet" (P18). As Engage allows for adding elements to the scene, it was soon "filled with animals, objects, special effects, and sounds" (P18). This was described as "absurdly funny" (P18) and "goofy" (P02) but during the meeting also as "distracting" (P17), and "less effective" (P01). Although the fun, interesting, and creative atmosphere in AltspaceVR, Engage and Gather Town can be appealing and support hedonic qualities of the system, it could be distracting from the subject of the meeting: "fun to use but not very helpful for serious work" (P11).

4.3.3 Social norms

Social interactions in VR depend on shared norms in a group as in reality. However, VEs enable behavior that is impossible in real life, such as walking through people, due to different physical or perceptual restrictions. The attendees shared experiences of how some social conventions translate well to virtual encounters: "I instinctively wanted to hug their avatar. When approaching, 1 m away, I noticed that this feels wrong. In reality, we would never hug each other [. . .], so I expressed my compassion verbally only" (P18). More often, participants reported novel social situations that might require new conventions. We identified two types of violating social norms: unintended impropriety and intended provocations.

Among the unintentional norm violations are situations in which, for example, attendees "felt bothered because the others came to close" (P16), or had unnatural postures due to the controllers sitting on the desk, causing awkwardness. One attendee described another uncomfortable situation due to technical challenges, in which he was concerned about having offended somebody with a joke as they did not react. It turned out that the other person was only staring at him quietly because they could not unmute to share the laughter. Other disruptions, however, were intended by users, such as an "eccentric avatar" (P18) that was seen as surprising at the professional occasion, somebody attending in a space suit, or users adding 3D models and effects to the scene. The most prominent example for this was "the monster which appeared in the middle of a student's presentation" (P17). The whipping daemon shown in Figure 9 caught the users' attention, distracted them from the presentation, and crashed the seriousness of the setting.

4.3.4 Lessons learned

The users appreciated the impression that their colleagues were in one room with them, especially when using an HMD. Barreda-Ángeles and Hartmann. (2022) found social presence to be a good predictor of relatedness and enjoyment, which is in line with our participants' reports. The high loading values of the overall meeting experience for the factors *F1 Involvement* and *F2 Co-Presence* indicate a strong link between the predictors. Analysis of qualitative data corroborates this as especially those attendees, who appreciated the high presence of others and feeling noticed by others in VR, were most fond of those meetings. Still, attendees rated their Zoom experiences regarding *F2 Co-Presence* higher than in VR. This finding is surprising considering the high immersion of VR and that many participants described the impression of being together in one room. However, the attendees of the investigated meetings are not strangers but colleagues, often for many years. Nevertheless, the avatars surrounding the user in VR seem unfamiliar and need a name tag to identify who it embodies. Without it, they could have been mistaken for intelligent game characters. In contrast, the identity of the conversation partners was unmistakable on video calls because of the visibility of familiar faces, which we discuss in the following subsection. Further, the ratings of the *F1 Involvement* factor were higher for video calls. From an autoethnographic perspective, we suspect that the meeting format has been adapted too tentatively to the medium. To exploit the advantages of interpersonal interactions in VR, we emphasize that the meeting purpose must benefit from high social presence, e.g., with team building or socializing goals. Not every format should be directly replicated virtually, as previous research supports (McVeigh-Schultz and Isbister, 2021b).

The discomfort of the attendees with the uncanny degree of avatar realism appeared to be more pronounced than the increase in social presence found by Zibrek and McDonnell. (2019) as a consequence of photorealism in social VR. In line with related work, the effects of visual fidelity of our test platforms' avatars seem more complex and depend on other factors, such as coherence, setting, and avatar behavior. We suggest focusing more on expressive and versatile avatar designs than sheer photorealism.

We can assume that some of the observed social inadequacies originated from technical overload or accidents. For example, the sense of personal space is not transferred directly from in-person

FIGURE 10
This figure shows the use of emojis to express emotions in the AltspaceVR platform.

interactions to VEs and depends on various factors, such as using an HMD or a desktop PC (Williamson et al., 2022). A prior training and familiarization phase with the system McVeigh-Schultz et al. (2018) is advisable. This may also reduce disruptions from users exploring features that the platform affords. We urge interaction designers to create systems that prevent users from awkward appearance or behavior by accident, e.g. while taking off equipment or sitting down and standing up. Further, until a common "VRtiquette" (Lahlou et al., 2021) is established over the years, the attendees should agree on shared rules and conventions that might be supported by regulations from the organizations–in the same way as video conferencing and hybrid meetings benefit from this (Saatçi et al., 2020).

4.4 Expression of emotions

The participants commented particularly much on the emotional aspects of interpersonal interactions and how replacing a video with an avatar changed their meeting experience. We must note that in the case of the team meetings investigated, most attendees typically turn on their cameras for the video call.

4.4.1 Facial Expressions and body Language

Seeing no facial expressions from the other attendees is one of the issues of VR meetings raised the most by our participants: "It is so odd to look at people and not see any reactions in their faces" (P12). Especially users who accessed the VE without an HMD and therefore lacked gesture and head tracking "looked a bit odd, less lively" to P12. The lack of faces was perceived as "disconcerting" (P07), "awkward" (P06), and "less direct, obtuse" (P10) compared to video calls. Therefore, interactions on video calls were described as "more expressive" (P14)—to "see people's faces and emotions" (P30) was important to connect for many. Several attendees have emphasized this concern throughout the experiment: "This is essential!" (P12). Above all, P02 missed the smiles and cats usually visible in the team's Zoom meetings. On the contrary, some attendees did not "need to see people's awkward or negative faces" (P13) and were relieved to be able to hide behind their avatars. They highlighted the advantages of missing cameras: "It looked more like a physical meeting because I can see everyone at once and not just the face. I could see people moving their heads and hands in VR" (P13). Several users addressed the expressiveness through avatar style and gestures: "I like the body language and the way people dress" (P08). One participant reflected on the impact that the different forms of encountering guests or strangers might have: "People present themselves differently. I even think you might get to know them differently in VR" (P12).

4.4.2 Emojis

To compensate for the lack of facial expressions, AltspaceVR allows spawning emojis above the user's avatar to react visually (see Figure 10). One attendee "expressed frustration with the –.– emoticons over her head" (P18) while having audio issues and no other means to communicate. A similar feature is also available on Zoom, despite other, more immediate ways to react: "When [the manager] asked the group to give quick reactions, most of the participants did not speak or gesture as a response but used the UI feature showing a thumbs-up or smiley" (P18). The participants explained that they missed this feature on Engage and StarLeaf. On Engage, they found the meetings "less 'affective' as I cannot show any emotion with the icons" (P13) as on AltspaceVR.

4.4.3 Lessons learned

Regardless of the primary purpose of meetings being of productive or social nature, the affective states of the attendees naturally have a critical impact on the collaboration. Whether for working or socializing with colleagues, it is essential to enable users to express their emotions and recognize those of others. This is especially pronounced with previously unknown people (Kirchner and Nordin Forsberg, 2021). Moustafa and Steed. (2018) found similar emotional states of users in social VR as in face-to-face settings that need to be considered. Still, our participants complained about the lack of facial expressions and other non-verbal communication, in line with the analysis by Tanenbaum et al. (2020). We conjecture that the missing faces are a significant reason for the lower co-presence ratings of VR compared to video calls in our exploratory factor analysis. Considering the study by Abdullah et al. (2021), this could also explain the participants' decreased activity to maintain social connectedness and less person-directed gaze in VR conversations compared to videoconferencing. With rapid progress in research on the recognition of facial expressions (Hamedi et al., 2018; Lou et al., 2020; Cha and Im, 2022) and eye movement (Schwartz et al., 2020), as well as their mapping to the avatar, more expressive face and gaze representation will be available in the near future. We expect this to be a crucial facilitator for authentic social exchange in VR. Until then, a rudimentary way of sharing one's emotions or reactions is needed for successful interaction in VR, such as emojis.

4.5 Meeting productivity

As information exchange was a major aim of the meetings, productivity was a central theme in the data. When participants compared the VR meetings with in-person meetings or video calls, they often referred to them as "less productive" (P05) and "inefficient, cumbersome" (P10). The reasons range from the inability to take notes and technical issues to the restrictions in seeing everyone simultaneously. Participants concluded that "the static meeting format" (P18) of "regular reports and presentations do not benefit from VR" (P05). However, there were also some positive comments on the productivity of meetings in VR, as people listened "with less distractions and multitasking" (P17) after a few AltspaceVR meetings.

4.5.1 Attention and distraction

Being in VR reduced distractions from other applications like mail, web browsers, or instant messaging for our participants. At the same time, it introduced new distractions to the meeting as people explored the features: "it can be distracting with all the non-meeting stuff you can do in VR" (P14). Overall there is no consensus on which medium is more beneficial regarding (in)attention, as some people expressed the need for a little diversion to focus on the meeting. Another aspect is the noticeable (in)attention of the meeting participants. Especially for the speaker, it is valuable feedback whether the audience actively listens. Some participants found that "in VR, one has a better awareness of the group, such as the closeness or distance, attention or distraction, *etc.*, which is missing in Zoom." (P25). In contrast, others thought "it was hard to tell [in VR] how distracted people were during the presentations [...] In the physical meetings, it is more obvious who's listening and who's busy with something else" (P18). P11 had the impression that on Zoom "the participants were more focused and could follow the conversation better."

4.5.2 Presentations

Giving and listening to presentations were reoccurring aspects. It felt for some participants more like an in-person setting. For the presenters, it was a more immersive and realistic experience than in video-based presentations: "I could walk in front of the screen and use my body. Felt like really presenting!" (P18). Similarly, P07 appreciated the "more true-to-life presentation environment with the projector screen and with the ability to see the audience spread out in a semicircle". Nevertheless, for the audience, problems known from in-person presentations, such as view-blocking or display resolution, reduced satisfaction. Attendees reported a better view of the slides in videoconferences.

4.5.3 Secondary tasks and tools

The access to notes had a substantial impact on the perceived productivity of the meeting, and comments like "I missed note taking" (P28) for VR were very common. Similarly, access to tools like the calendar, search engines, and instant messengers for discreet parallel exchange within the team was missed. The lack of simultaneous messaging was mentioned frequently as it led to "less interactive" (P17) meetings. In videoconferences, "most interactions occurred 'outside the meeting'" (P28) simultaneously *via* instant messaging. Several attendees mentioned that they could not do other productive tasks while being "trapped in VR" (P28), such as answering emails during personally irrelevant parts of the meeting. Also, the inability to access physical objects as beverages during VR sessions was criticized: "I miss my tea" (P13) and "my coffee got cold" (P18).

4.5.4 Lessons learned

For our work-focused meetings, the participants greatly appreciated the possibility of taking notes and accessing additional information. These aspects are not yet adequately provided by the VR tools tested in the scope of this study or require an unreasonable amount of individual effort. The accustomed workflow of a user should be supported in virtual meeting environments without allowing too many distracting elements into the VE. Recent developments promise a smoother integration of physical keyboards and applications from the desktop environment (Otte et al., 2019; Oculus, 2021), which could help to bridge VEs and the usual workspace of a user.

Although Sarkar et al. (2021) found parallel chatting in video calls distracting and overwhelming for some users, they also identified valuable benefits from it that resonate with our participants' responses. More research needs to be done on integrating discreet side communication in VR. On the other hand, compared to video conferencing tools, participants valued the high-fidelity and engaging way of presenting in VR. This close resemblance matches the findings by Ginkel et al. (2019) who measured similar skill improvement in an oral presentation training in VR as in a physical environment. But as similarly shown by Yoshimura and Borst. (2021), the satisfaction with VR presentations also depended for our participants on the (dis) comfort with the equipment. Short and interactive talks can benefit more from VR than long or many presentations.

4.6 User needs

The primary objective of the attendees in our meetings was to get together once a week and exchange information. Their individual needs for functionality, comfort, and appropriate self-representation must be met.

4.6.1 Technical Literacy of users

The questionnaire answers revealed some aspects rooted in the group's diversity. Although most of the participants had considerable knowledge of technology in general, some had serious technical challenges that sometimes even prevented them from participating at all, "because I was not able to set up the app" (P19), or required considerable effort: "I needed nearly an hour to join the meeting (and then just *via* [desktop PC]). And on top of that, I later needed help to get the VR setup running." (P25). Even for experienced users, it took time to get familiar with the system as they encountered initial issues with the setup, updates, or controls. This clearly shows that even when using standard devices and commercially available applications, VR meetings still pose challenges to people who do not use the technology regularly. P09 was concerned about guests because "it is more of a hassle as they have to get accustomed to the software as well." In particular, P23 worried about students presenting their thesis progress as it "could add to the pressure that they are already under and make them even more uncomfortable and stressed".

4.6.2 HMD discomfort

One frequently mentioned theme was the discomfort related to the VR headset. The participants reported motion sickness, headaches, that the "headset felt heavy and uncomfortable after some time in the meeting" (P15), and that it was "a bit stressful for my eyes" (P12). Participants used different HMD models, and some were unsatisfied with the visual quality. For example, P12 said about AltspaceVR: "the slides were difficult to read, the letters flickering; this was uncomfortable". Furthermore, P17 mentioned that the Oculus Quest "got quite laggy with so many people. I had to reset the graphic settings. It gives me a headache and I'm more exhausted afterward." But using the desktop client was also not a good alternative for some participants as they reported that "using 2D VR is a bit inconvenient for the control. It is somehow like [a first-person shooter], but I am not very used to that." (P13)

4.6.3 Privacy

Some people described that they "felt observed" (P12) on video calls because everybody could "see me and my living room" (P18). Regarding personal privacy and discretion, the VR sessions received positive feedback, as they allowed users to be invisible behind the avatar. Therefore, they did not need to worry about a presentable appearance or "need to show [their] facial expressions" (P13). Some participants also felt more confident as it "kinda helps with anxiety" (P14). However, users also reported that the spatial audio settings in AltspaceVR were insufficient for private conversations. P09 felt like "intruding" in private conversations: "After the meeting, it was weird as people talked in smaller groups, but one could hear everything".

4.6.4 Lessons learned

While some appreciated the concealment of personal appearance, emotional reactions, and the home office environment, private conversations could be heard from far away. After the effects of Zoom fatigue (Shockley et al., 2021) during the pandemic, our participants felt relieved from the stress of being observed. Although the quantitative results did not show any differences between the platforms in terms of factor *F3 Privacy*, privacy was of concern in the qualitative data. The reason for this discrepancy is the different interpretations of the term. While standardized questionnaires and legal frameworks often consider privacy as the protection of information from the service provider, our participants were more concerned about the other attendees invading their privacy or being visually exposed to others. Quantitative methods should consider aspects of the personal privacy of meeting attendees beyond data protection regulations.

The fact that attendees from an HCI research group struggled with technical barriers clearly shows that meeting platforms must cater to users' diverse needs and abilities. Usability issues in many established social VR platforms, including AltspaceVR, were also identified by Liu and Steed. (2021). Using current off-the-shelf VR hardware and software resulted in a stressful experience with technical issues and HMD discomfort for some participants. Until the onboarding experience is smooth and quick for everyone, organizers must assume responsibility for providing support before the first meeting, as recommended similarly by Lahlou et al. (2021). It seems that VR meetings are currently not as inclusive and accessible as video-based platforms due to a higher technical entry barrier, even if expensive equipment is provided.

4.7 Platform choice for future meetings

In the final survey after the last meeting, we asked participants to draw a conclusion comparing the different platforms and to choose a long-term preference. Many participants enjoyed exploring the possibility of meeting the team in VR: "it sure was fun and showed the potential for the future" (P09). The users appreciated the possibilities of "free spatial arrangement" (P12) with benefits especially for "engaging conversations between multiple actors" (P05) as it allows "spontaneously splitting into groups and chatting in clusters, without losing the overview of who else is present in the global room" (P21). Nevertheless, only two survey participants selected a VR platform as their preferred meeting medium. P14 appreciated that it "is more fun and kinda helps with anxiety". P18 supported using VR for future meetings because "on Zoom, it always feels like people cannot wait to get out again. Most participants try to be as quiet as possible not to delay the meeting unnecessarily. In VR, there was much more socializing, arriving early and staying longer, personal and informal exchange, having fun, or enjoying meeting with others."

However, the vast majority argued for returning to desktop-based video conferencing tools. Of the 15 participants who completed the survey after the last session, 13 preferred to go back to Zoom or StarLeaf. For most attendees, the "regular reports and presentations do not benefit from VR" (P05) nor from the spatial component of Gather Town. Although fun, VEs were perceived as "not very helpful for serious work" (P11). P05 criticized the attempt "to replicate the reality without adding specific value that comes from the medium." On the contrary, for VR, the preparations and technical issues were disproportionately time-consuming, caused motion sickness for some attendees, secondary tasks were unavailable, and it was repeatedly described as inconvenient. For P19, the experience "was like speaking on an old phone with multiple participants at the same time, wearing a rock tied to my head, and watching some meaningless cartoon simultaneously". On the other hand, in video calls, it is possible to see the faces of "people, and it works best. No preparations. No nausea or headache" (P02). It was described as "most convenient" (P09) and "more productive" (P13) with easy "access for external visitors" (P11). As many others, P12 "learned that seeing the faces of others is so important."

5 Limitations and future work

We conducted this study in the wild during the regular team meetings of the research lab. Therefore, it has some limitations. Following an autoethnographic approach, the authors are part of the team and participated in the study. However, we ensured that the anonymization remained intact during coding and analysis. The study sample represents a very heterogeneous group as some team members have high technical literacy, and administrative staff was involved. This is most likely not representative of the general population, but for the domains that will be early adopters, this should be close to a representative sample. We also point out that most meeting attendees had no previous experience holding meetings in VR. Hence, the findings apply to teams of novice users and should only be extended to experts with caution. In addition, there was a trade-off between the length of the questionnaire and the aspects to cover, which influenced the direction of the results.

To keep the atmosphere and behavior of the participants as natural as possible, we did not record the sessions. Consequently, no systematic observations of the attendees' behavior were possible, and we can only report on aspects described in the questionnaires. The selection, order, and frequency of the used platforms were team decisions independent of the study objectives, and therefore, the platforms were not explored evenly. Additionally, there might be possible biasing effects for the VR platform used second (Engage), as most users now had previous experiences with social VR. Furthermore, the content, length, and need for follow-up conversations varied every week due to circumstances, making it challenging to compare the meetings directly. During the meetings, various system configurations were used, which created different experiences for the individual users. We had a variety of HMDs (HTC Vive (Pro), Valve Index, and Oculus Quest), operating systems (Mac, Windows, Linux), and internet browsers (relevant for Gather Town). While this limits comparability between users, this case study was done in an authentic context with meetings that would have taken place in this manner regardless. All this ensures an authentic setting, leading to high external validity.

Based on the outcome of this study, further research is needed with a clear separation of researchers and meeting attendees. At the cost of the insights from an autoethnographic approach, the generalizability can be higher with controlled, balanced, and diverse samples across domains and prior experience. Moreover, as a result of the independence of the meeting participants from the researchers, no negative consequences of honest answers are to be expected, which allows to lift the anonymity towards the data analysts and thus extends methodological possibilities. With a team that accepts the given parameters of the experiment, such as the choice of platforms and hardware or video recording of the meetings, future work could rely on more systematic and controlled circumstances enabling behavioral and linguistic analyses with high comparability. Still, a careful balance between external provisions and internal authenticity must be achieved.

6 Conclusion

This in-the-wild case study on using social VR platforms and video conferencing software for weekly team meetings over the course of 4 months provides authentic insights into the experiences of 32 participants. Through questionnaires, we collected responses to open questions and Likert scale items after every meeting on various topics related to social interaction, meeting productivity, and individual experiences. In a thematic analysis of the qualitative data, we found five prominent themes that cover spatial aspects, meeting atmosphere, expression of emotions, meeting productivity, and user needs. The results show that regarding realistic group dynamics in gatherings before and after the official meeting, or for the impression of being together in the same place, VR meetings resemble in-person meetings more than videoconferences. Nevertheless, at the same time, videoconferences provided a closer-to-reality experience regarding seeing others' real faces and emotions, discreet side communication, and support of secondary tasks than VR meetings. Furthermore, we performed a factor analysis on the quantitative data, revealing the advantages of videoconferences over VR platforms regarding ratings on involvement and co-presence. This finding is in line with the assessments in the final survey, in which participants were asked to draw a conclusion comparing the different platforms. Here, only two out of 15 survey participants selected a VR platform as their preferred meeting medium for weekly meetings in the future. Not being able to see the real faces of their colleagues was one of the most decisive factors for many participants. Therefore, at the end of the case study, the team decided to return to videoconferences for the long term. Some aspects found in this study have been discussed in related literature before. Still, this authentic in-the-wild setting of intrinsically motivated exploration of how suitable current commercial platforms are under natural working conditions enabled us to bring practical issues into context and highlight critical research gaps.

Our results indicate that the currently available off-the-shelf social VR platforms tested in this study do not yet sufficiently meet the needs and preferences of users to attend weekly team meetings like ours. Although they provide advantages in certain aspects, limitations such as time-consuming preparations, technical issues, HMD discomfort, motion sickness, and unavailable secondary tasks were among the main reasons for returning to video conferencing. Moreover, the findings indicate that the meeting format was not ideal for fully using the benefits of social VR. Above all, our results reveal how important it is for meeting attendees to see their colleagues' facial expressions and emotions and the current limitations of VR platforms in this regard. We suspect this to be the strongest contributor to the higher co-presence ratings of video calls compared to VR that we unexpectedly measured.

Some of the problems we encountered on our journey might be resolved within this decade thanks to technical advancements. As outlined in the Lessons Learned subsections above, new technologies using electromyogram and neuromuscular signals, as well as eye gaze and facial reconstruction, will provide avatars with rich emotional expressivity. Also, body language will be more realistic with advanced sensors and inverse kinematics. Peripherals such as physical keyboards and other components or applications of existing workflows will be seamlessly integrated into the virtual meeting environment. This considerably improves text input capabilities and the support of secondary tasks. In the meantime, usability issues on the software side will be resolved, and the devices will become lighter and more convenient. Prospectively, if this case study had been conducted in a few years, the outcome concerning these challenges might be completely different.

However, other challenges we encountered in our study might not be possible to overcome with technical progress and require careful consideration when (1) preparing technical aspects of the VE, and (2) planning and moderating the meeting. Considering our sample, these insights apply particularly to users who are new to having meetings in VR. First, as part of the technical setup, an onboarding tutorial before the meeting must ensure that all attendants are familiar with the application and that their system is up-to-date. Users with previous experience should have had the opportunity to explore features in the VE before the meeting starts so it would not distract them. Further, the spatial audio should be configured for plenum situations in which speakers are audible by everyone anywhere but also cater for private (group) conversations with privacy-protecting sound bubbles and little distraction from distant background conversations. Acoustic parameters must always be transparent to users. The application's overall visual style and professional appearance should match the seriousness of the occasion and the attendees' preferences. We recommend disabling permissions to insert distracting objects or special effects to the scene when not required for the meeting goal.

Second, to adapt the meeting format to the platform, a code of conduct should be agreed on to ensure a comfortable environment for everyone. Among other conventions, it should address proxemics, muting, recognizability, and modifications of the shared scene. In line with related work, we found it important to designate a person responsible for technical support and facilitation, such as amplifying or muting attendees. Generally, the format and goals of a meeting should be adapted to the advantages of social encounters in VR. Ideally, the reason for gathering is interpersonal interaction in dynamic group constellations with rich opportunities for socialization and creative exchange. Our participants appeared to appreciate situations before and after the meetings that afford the CoFIRe steps proposed by Erickson et al. (2011): **co**alescence into small groups, **f**ocused **i**nteractions without distractions, and eventually **re**mixing with others. We recommend making use of spatial advantages and meaningful movements, such as forming queues, circles, and groups, flying when appropriate, moderating and taking turns with gaze and gestures, as well as organizing meeting content systematically within the space as demonstrated by Luo et al. (2022) for augmented reality.

We learned most of these lessons only as the study progressed. Initially, videoconferences were a substitute for in-person meetings due to the COVID-19 pandemic. The VR platforms were experimental substitutes for the new *status quo* of video conferencing. The reason for using any of these systems, however, was not to replicate in-person meetings but to achieve the purpose of the meetings. How the group aimed to achieve these goals was organically shaped within the conditions of the respective platforms. Nevertheless, the meeting format was not sufficiently adapted to the benefits and restrictions of social VR platforms. Technology does not break habits. And they appear to have been strong from the influence of a traditional meeting concept manifested over many years. The adjustment on-the-fly was not sufficient. Instead, a more conscious and targeted adaptation would have been required. Similarly, McVeigh-Schultz and Isbister. (2021b) argue for using XR to deliberately deviate from direct replication of face-to-face meetings, such as with augmentation of social behavior (Roth et al., 2018; Roth et al., 2019) or "social superpowers" (McVeigh-Schultz and Isbister, 2021a).

A potential direction for the future could be a seamless integration of the different platform types to allow dynamic rearrangement according to the meeting situation, which often changes depending on the ongoing activities. This would allow using the most powerful medium for each case. For example, interactivity, small group interaction, poster sessions, or social events could take place in immersive XR environments, and detailed presentations or one-to-many announcements through video conferencing channels–with immediate transition in between. Similar to the hybrid concept of Gather Town, this proposal attempts to combine the best of two solutions, although not limited to one device or a 2D world.

Overall, our case study provides authentic insights into conducting team meetings on off-the-shelf virtual reality platforms that can inform the appropriate choices and configurations of the platform, adaptations of the meeting format, and future requirements of social VR platforms.

Data availability statement

The quantitative dataset presented in this article can be found at\url (https://doi.org/10.17605/OSF.IO/YG7HK). The qualitative dataset presented in this article is not publicly available because the participants in the case study can easily be identified by their affiliation. Therefore, we assure them of the complete confidentiality of the collected data. Only the authors, excluding the group manager, could access the data to protect participants' privacy and encourage them to respond authentically without fear of repercussions.

Ethics statement

The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

All authors developed the study design. MB prepared the test environments together with AR and coordinated the project. MB, AR, SP, YL, DA, RM, and TD carried out the data acquisition after the meetings. MB, AR, SP, YL, and TD were mainly responsible for data analysis. MB, AR, SP, and TD were responsible for drafting the manuscript. MB, AR, SP, and TD revised and finalized the manuscript.

Funding

This research was partially funded by the Klaus Tschira Stiftung by financing the scholarship of MB. The open access publication fees were covered by Staats-und Universitätsbibliothek Bremen.

Acknowledgments

We want to thank all team members for their commitment, feedback, and patience in this study. Also, we want to thank the Engage team, who provided us with a free trial for larger groups to run our study. A preprint of this paper is available on arXiv.org (Bonfert et al., 2022).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frvir.2022.945791/full#supplementary-material

References

Abdullah, A., Kolkmeier, J., Lo, V., and Neff, M. (2021). Videoconference and embodied VR: Communication patterns across task and medium. *Proc. ACM Human-Computer Interact.* 5, 129–29. doi:10.1145/3479597

Ahn, S. J. G., Levy, L., Eden, A., Won, A. S., MacIntyre, B., and Johnsen, K. (2021). IEEEVR2020: Exploring the first steps toward standalone virtual conferences. *Front. Virtual Real* 2, 648575. doi:10.3389/frvir.2021.648575

Bailenson, J. N., Beall, A. C., Loomis, J., Blascovich, J., and Turk, M. (2004). Transformed social interaction: Decoupling representation from behavior and form in collaborative virtual environments. *Presence Teleoperators Virtual Environ.* 13, 428–441. doi:10.1162/1054746041944803

Bailenson, J. N., Blascovich, J., Beall, A. C., and Loomis, J. M. (2001). Equilibrium theory revisited: Mutual gaze and personal space in virtual environments. *Presence Teleoperators Virtual Environ.* 10, 583–598. doi:10.1162/105474601753272844

Bailenson, J. N. (2021). Nonverbal overload: A theoretical argument for the causes of Zoom fatigue. *Technol. Mind, Behav.* 2. doi:10.1037/tmb0000030

Bailenson, J. N., Swinth, K., Hoyt, C., Persky, S., Dimov, A., and Blascovich, J. (2005). The independent and interactive effects of embodied-agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments. *Presence Teleoperators Virtual Environ.* 14, 379–393. doi:10.1162/105474605774785235

Barreda-Ángeles, M., and Hartmann, T. (2022). Psychological benefits of using social virtual reality platforms during the Covid-19 pandemic: The role of social and spatial presence. *Comput. Hum. Behav.* 127, 107047. doi:10.1016/j.chb.2021.107047

Bente, G., Rüggenberg, S., Krämer, N. C., and Eschenburg, F. (2008). Avatar-mediated networking: Increasing social presence and interpersonal trust in net-based collaborations. *Hum. Commun. Res.* 34, 287–318. doi:10.1111/j.1468-2958.2008.00322.x

Blackwell, L., Ellison, N., Elliott-Deflo, N., and Schwartz, R. (2019b). Harassment in social virtual reality: Challenges for platform governance. *Proc. ACM Human-Computer Interact.* 3, 1125–25. doi:10.1145/3359202

Blackwell, L., Ellison, N., Elliott-Deflo, N., and Schwartz, R. (2019a). "Harassment in social VR: Implications for design," in 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Osaka, Japan, 23-27 March 2019 (IEEE), 854–855. doi:10.1109/VR.2019.8798165

Bleakley, A., Rough, D., Edwards, J., Doyle, P., Dumbleton, O., Clark, L., et al. (2022). Bridging social distance during social distancing: Exploring social talk and remote collegiality in video conferencing. *Human–Computer Interact.* 37, 404–432. doi:10.1080/07370024.2021.1994859

Bonfert, M., Reinschluessel, A. V., Putze, S., Lai, Y., Alexandrovsky, D., Malaka, R., et al. (2022). Seeing the faces is so important" – experiences from online team meetings on commercial virtual reality platforms. arXiv. doi:10.48550/arXiv.2210

Brucks, M. S., and Levav, J. (2022). Virtual communication curbs creative idea generation. *Nature* 605, 108–112. doi:10.1038/s41586-022-04643-y

Burgoon, J. K., Bonito, J. A., Bengtsson, B., Ramirez, A., Dunbar, N. E., and Miczo, N. (1999). Testing the interactivity model: Communication processes, partner assessments, and the quality of collaborative work. *J. Manag. Inf. Syst.* 16, 33–56. doi:10.1080/07421222.1999.11518255

Byun, B., Awasthi, A., Chou, P. A., Kapoor, A., Lee, B., and Czerwinski, M. (2011). "Honest signals in video conferencing," in 2011 IEEE International Conference on Multimedia and Expo, Barcelona, Spain, 11-15 July 2011 (IEEE), 1–6. doi:10.1109/ICME.2011.6011855

Cha, H. S., and Im, C. H. (2022). Performance enhancement of facial electromyogram-based facial-expression recognition for social virtual reality applications using linear discriminant analysis adaptation. *Virtual Real.* 26, 385–398. doi:10.1007/s10055-021-00575-6

Daft, R. L., and Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *Manag. Sci.* 32, 554–571. doi:10.1287/mnsc.32.5.554

Dzardanova, E., Kasapakis, V., Gavalas, D., and Sylaiou, S. (2021). Virtual reality as a communication medium: A comparative study of forced compliance in virtual reality versus physical world. *Virtual Real* 26, 737–757. doi:10.1007/s10055-021-00564-9

Erickson, T., Shami, N. S., Kellogg, W. A., and Levine, D. W. (2011). "Synchronous interaction among hundreds: An evaluation of a conference in an avatar-based virtual environment," in *Proceedings of the SIGCHI conference on human factors in computing systems* (New York, NY, USA: Association for Computing Machinery), 503–512. doi:10.1145/1978942.1979013

Fägersten, K. (2010). "Using discourse analysis to assess social co-presence in the video conference environment," in *Cases on online discussion and interaction: Experiences and outcomes* (Hershey, PA: IGI Global), 19.

Finn, K. E., Sellen, A. J., and Wilbur, S. (1997). *Video-mediated communication*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Freeman, G., Zamanifard, S., Maloney, D., and Acena, D. (2022). Disturbing the peace: Experiencing and mitigating emerging harassment in social virtual reality. *Proc. ACM Hum. Comput. Interact.* 6, 1–30. doi:10.1145/3512932

Galegher, J., Kraut, R. E., and Egido, C. (2013). *Intellectual teamwork: Social and technological foundations of cooperative work*. New York: Psychology Press. doi:10.4324/9781315807645

Ginkel, S. v., Gulikers, J., Biemans, H., Noroozi, O., Roozen, M., Bos, T., et al. (2019). Fostering oral presentation competence through a virtual reality-based task for delivering feedback. *Comput. Educ.* 134, 78–97. doi:10.1016/j.compedu.2019.02.006

Gonzalez-Franco, M., Maselli, A., Florencio, D., Smolyanskiy, N., and Zhang, Z. (2017). Concurrent talking in immersive virtual reality: On the dominance of visual speech cues. *Sci. Rep.* 7, 3817. doi:10.1038/s41598-017-04201-x

Hamedi, M., Salleh, S. H., Ting, C. M., Astaraki, M., and Noor, A. M. (2018). Robust facial expression recognition for MuCI: A comprehensive neuromuscular signal analysis. *IEEE Trans. Affect. Comput.* 9, 102–115. doi:10.1109/TAFFC.2016.2569098

Hecht, H., Welsch, R., Viehoff, J., and Longo, M. R. (2019). The shape of personal space. *Acta Psychol.* 193, 113–122. doi:10.1016/j.actpsy.2018.12.009

Hinds, P. J. (1999). The cognitive and interpersonal costs of video. *Media Psychol.* 1, 283–311. doi:10.1207/s1532785xmep0104_1

Kang, S. H., and Watt, J. H. (2013). The impact of avatar realism and anonymity on effective communication via mobile devices. *Comput. Hum. Behav.* 29, 1169–1181. doi:10.1016/j.chb.2012.10.010

Kern, A. C., and Ellermeier, W. (2020). Audio in VR: Effects of a soundscape and movement-triggered step sounds on presence. *Front. Robotics AI* 7, 20. doi:10.3389/frobt.2020.00020

Kirchner, K., and Nordin Forsberg, B. (2021). "A conference goes virtual: Lessons from creating a social event in the virtual reality," in *Innovations for community services*. Editors U. R. Krieger, G. Eichler, C. Erfurth, and G. Fahrnberger (Cham: Springer International Publishing), 123–134. doi:10.1007/978-3-030-75004-6_9

Koseki, N., Iijima, M., Yamamoto, Y., and Seiki, K. (2020). "Identifying communication problems in the use of video conferencing systems," in 2020 9th International Congress on Advanced Applied Informatics (IIAI-AAI), Kitakyushu, Japan, 01-15 September 2020 (IEEE), 651–658. doi:10.1109/IIAI-AAI50415.2020.00133

Kramer, A. D. I., Oh, L. M., and Fussell, S. R. (2006). "Using linguistic features to measure presence in computer-mediated communication," in *Proceedings of the SIGCHI conference on human factors in computing systems* (New York, NY, USA: Association for Computing Machinery), 913–916. doi:10.1145/1124772.1124907

Kyrlitsias, C., and Michael-Grigoriou, D. (2022). Social interaction with agents and avatars in immersive virtual environments. *A Surv.* 2, 786665. doi:10.3389/frvir.2021.786665

Lahlou, S., Pea, R., Heitmayer, M., Russell, G. M., Schimmelpfennig, R., Yamin, P., et al. (2021). "Are we 'Beyond being there' yet?: Towards better interweaving epistemic and social aspects of virtual reality conferencing," in *Extended abstracts of the 2021 CHI conference on human factors in computing systems* (Yokohama Japan: ACM), 1–6. doi:10.1145/3411763.3451579.00000

Liu, Q., and Steed, A. (2021). Social virtual reality platform comparison and evaluation using a guided group walkthrough method. *Front. Virtual Real.* 2, 668181. doi:10.3389/frvir.2021.668181

Lou, J., Wang, Y., Nduka, C., Hamedi, M., Mavridou, I., Wang, F. Y., et al. (2020). Realistic facial expression reconstruction for VR HMD users. *IEEE Trans. Multimedia* 22, 730–743. doi:10.1109/TMM.2019.2933338

Luo, W., Lehmann, A., Widengren, H., and Dachselt, R. (2022). "Where should we put it? Layout and placement strategies of documents in augmented reality for collaborative sensemaking," in *CHI conference on human factors in computing systems* (New York, NY, USA: Association for Computing Machinery), 1–16. doi:10.1145/3491102.3501946

Mack, K., Das, M., Jain, D., Bragg, D., Tang, J., Begel, A., et al. (2021). "Mixed abilities and varied experiences: A group autoethnography of a virtual summer internship," in *The 23rd international ACM SIGACCESS conference on computers and accessibility* (New York, NY, USA: Association for Computing Machinery), 1–13. doi:10.1145/3441852.3471199

Maloney, D., Freeman, G., and Wohn, D. Y. (2020). Talking without a voice. *Talk. without a voice" Underst. non-verbal Commun. Soc. virtual Real.* 4, 1125–25. doi:10.1145/3415246

McVeigh-Schultz, J., and Isbister, K. (2021b). A "beyond being there" for VR meetings: Envisioning the future of remote work. *Human–Computer Interact.* 37, 433–453. doi:10.1080/07370024.2021.1994860

McVeigh-Schultz, J., and Isbister, K. (2021a). "The case for "weird social" in VR/XR: A vision of social superpowers beyond meatspace," in *Extended abstracts of the 2021 CHI conference on human factors in computing systems* (New York, NY, USA: Association for Computing Machinery), 1–10.

McVeigh-Schultz, J., Kolesnichenko, A., and Isbister, K. (2019). "Shaping pro-social interaction in VR: An emerging design framework," in *Proceedings of the 2019 CHI conference on human factors in computing systems* (New York, NY, USA: Association for Computing Machinery). doi:10.1145/3290605.3300794

McVeigh-Schultz, J., Márquez Segura, E., Merrill, N., and Isbister, K. (2018). "What's it mean to "Be social" in VR? Mapping the social VR design ecology," in *Proceedings of the 2018 ACM conference companion publication on designing interactive systems* (New York, NY, USA: Association for Computing Machinery), DIS '18 Companion), 289–294. doi:10.1145/3197391.3205451

Microsoft (2022). Work trend index 2022. great expectations: Making hybrid work work. Annual Report.

Milgram, S., and Gudehus, C. (1978). *Obedience to authority*. Harper & Row.

Moustafa, F., and Steed, A. (2018). "A longitudinal study of small group interaction in social virtual reality," in *Proceedings of the 24th ACM symposium on virtual reality software and technology* (New York, NY, USA: Association for Computing Machinery), 1–10. doi:10.1145/3281505.3281527

Nadler, R. (2020). Understanding "Zoom fatigue": Theorizing spatial dynamics as third skins in computer-mediated communication. *Comput. Compos.* 58, 102613. doi:10.1016/j.compcom.2020.102613

Nilles, J. (1975). Telecommunications and organizational decentralization. *IEEE Trans. Commun.* 23, 1142–1147. doi:10.1109/TCOM.1975.1092687

Nind, M., Coverdale, A., and Meckin, R. (2021). Changing social research practices in the context of covid-19: Rapid evidence review. Working Paper, NCRM. doi:10.5258/NCRM/NCRM.00004458

Nowak, K. (2001). "Defining and differentiating copresence, social presence and presence as transportation," in *4 th annual international workshop* (Philadelphia: Lawrence Erlbaum Associates).

Nowak, K. L., and Biocca, F. (2003). The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence Teleoperators Virtual Environ.* 12, 481–494. doi:10.1162/105474603322761289

Ochsman, R. B., and Chapanis, A. (1974). The effects of 10 communication modes on the behavior of teams during Co-operative problem-solving. *Int. J. Man. Mach. Stud.* 6, 579–619. doi:10.1016/S0020-7373(74)80019-2

Oculus (2021). *Introducing 'Horizon Workrooms': Remote collaboration reimagined*. Menlo Park, CA: Meta Platforms Inc.

OECD (2021). "Teleworking in the COVID-19 pandemic: Trends and prospects," in *Tech. rep.* (Paris, France: Organisation for Economic Co-operation and Development).

Oh, C. S., Bailenson, J. N., and Welch, G. F. (2018). A systematic review of social presence: Definition, antecedents, and implications. *Front. Robotics AI* 5, 114. doi:10.3389/frobt.2018.00114

Olson, J. S., Olson, G. M., and Meader, D. (1997). "Face-to-face group work compared to remote group work with and without video," in *Video-mediated communication* (Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers).

Otte, A., Menzner, T., Gesslein, T., Gagel, P., Schneider, D., and Grubert, J. (2019). "Towards utilizing touch-sensitive physical keyboards for text entry in virtual reality," in 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Osaka, Japan, 23-27 March 2019 (IEEE), 1729–1732. doi:10.1109/VR.2019.8797740

Poeschl, S., Wall, K., and Doering, N. (2013). "Integration of spatial sound in immersive virtual environments an experimental study on effects of spatial sound on presence," in 2013 IEEE Virtual Reality (VR), Lake Buena Vista, FL, USA, 18-20 March 2013 (IEEE). doi:10.1109/VR.2013.6549396

Raghuram, S., Hill, N. S., Gibbs, J. L., and Maruping, L. M. (2019). Virtual work: Bridging research clusters. *Acad. Manag. Ann.* 13, 308–341. doi:10.5465/annals.2017.0020

Rogers, B., Masoodian, M., and Apperley, M. (2018). "A virtual cocktail party: Supporting informal social interactions in a virtual conference," in *Proceedings of the 2018 international conference on advanced visual interfaces* (New York, NY, USA: Association for Computing Machinery), 1–3. doi:10.1145/3206505.3206569

Roth, D., Bente, G., Kullmann, P., Mal, D., Purps, C. F., Vogeley, K., et al. (2019). "Technologies for social augmentations in user-embodied virtual reality," in *25th ACM symposium on virtual reality software and technology* (New York, NY, USA: Association for Computing Machinery), 1–12. doi:10.1145/3359996.3364269

Roth, D., Klelnbeck, C., Feigl, T., Mutschler, C., and Latoschik, M. E. (2018). "Beyond replication: Augmenting social behaviors in multi-user virtual realities," in 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Tuebingen/Reutlingen, Germany, 18-22 March 2018 (IEEE), 215–222. doi:10.1109/VR.2018.8447550

Rzeszewski, M., and Evans, L. (2020). Virtual place during quarantine – A curious case of VRChat. *Rozw. Reg. i Polityka Reg.* 57, 57–75. doi:10.14746/rrpr.2020.51.06

Saatçi, B., Akyüz, K., Rintel, S., and Klokmose, C. N. (2020). Re)Configuring hybrid meetings: Moving from user-centered design to meeting-centered design. *Comput. Support. Coop. Work (CSCW)* 29, 769–794. doi:10.1007/s10606-020-09385-x

Samrose, S., McDuff, D., Sim, R., Suh, J., Rowan, K., Hernandez, J., et al. (2021). "Meetingcoach: An intelligent dashboard for supporting effective & inclusive meetings," in *Proceedings of the 2021 CHI conference on human factors in computing systems* (New York, NY, USA: Association for Computing Machinery). doi:10.1145/3411764.3445615

Samrose, S., Zhao, R., White, J., Li, V., Nova, L., Lu, Y., et al. (2018). Coco: Collaboration coach for understanding team dynamics during video conferencing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 1–24. doi:10.1145/3161186

Sarkar, A., Rintel, S., Borowiec, D., Bergmann, R., Gillett, S., Bragg, D., et al. (2021). "The promise and peril of parallel chat in video meetings for work," in *Extended abstracts of the 2021 CHI conference on human factors in computing systems* (New York, NY, USA: Association for Computing Machinery). doi:10.1145/3411763.3451793.00003

Scavarelli, A., Arya, A., and Teather, R. J. (2021). Virtual reality and augmented reality in social learning spaces: A literature review. *Virtual Real.* 25, 257–277. doi:10.1007/s10055-020-00444-8

Schwartz, G., Wei, S. E., Wang, T. L., Lombardi, S., Simon, T., Saragih, J., et al. (2020). The eyes have it: An integrated eye and face model for photorealistic facial animation. *ACM Trans. Graph.* 39, 91. doi:10.1145/3386569.3392493

Shockley, K. M., Gabriel, A. S., Robertson, D., Rosen, C. C., Chawla, N., Ganster, M. L., et al. (2021). The fatiguing effects of camera use in virtual meetings: A within-person field experiment. *J. Appl. Psychol.* 106, 1137–1155. doi:10.1037/apl0000948

Short, J., Williams, E., and Christie, B. (1976). *The social psychology of telecommunications*. Toronto; London; New York: Wiley.

Slater, M., and Sanchez-Vives, M. V. (2016). Enhancing our lives with immersive virtual reality. *Front. Robot.* 3, 74. doi:10.3389/frobt.2016.00074

Ståhl, O. (1999). "Meetings for real–experiences from a series of vr-based project meetings," in *Proceedings of the ACM symposium on virtual reality software and technology* (New York, NY, USA: Association for Computing Machinery), 164–165. doi:10.1145/323663.323691

Suh, H., Shahriaree, N., Hekler, E. B., and Kientz, J. A. (2016). "Developing and validating the user burden scale: A tool for assessing user burden in computing systems," in *Proceedings of the 2016 CHI conference on human factors in computing systems* (New York, NY, USA: Association for Computing Machinery), 3988–3999.

Sykownik, P., Graf, L., Zils, C., and Masuch, M. (2021). "The most social platform ever? A survey about activities amp; motives of social VR users," in 2021 IEEE Virtual Reality and 3D User Interfaces (VR), Lisboa, Portugal, 27 March 2021 - 01 April 2021 (IEEE). doi:10.1109/VR50410.2021

Tanenbaum, T. J., Hartoonian, N., and Bryan, J. (2020). "How do I make this thing smile?": An inventory of expressive nonverbal communication in commercial social virtual reality platforms," in *Proceedings of the 2020 CHI conference on human factors in computing systems* (Honolulu HI USA: ACM), 1–13. doi:10.1145/3313831.3376606

Torro, O., Jalo, H., and Pirkkalainen, H. (2021). Six reasons why virtual reality is a game-changing computing and communication platform for organizations. *Commun. ACM* 64, 48–55. doi:10.1145/3440868

Wigham, C. R., and Chanier, T. (2013). A study of verbal and nonverbal communication in second life – The ARCHI21 experience. *ReCALL* 25, 63–84. doi:10.1017/S0958344012000250

Wilcox, L. M., Allison, R. S., Elfassy, S., and Grelik, C. (2006). Personal space in virtual reality. *ACM Trans. Appl. Percept.* 3, 412–428. doi:10.1145/1190036.1190041

Williamson, J., Li, J., Vinayagamoorthy, V., Shamma, D. A., and Cesar, P. (2021). "Proxemics and social interactions in an instrumented virtual reality workshop," in *Proceedings of the 2021 CHI conference on human factors in computing systems* (New York, NY, USA: Association for Computing Machinery). doi:10.1145/3411764.3445729

Williamson, J. R., O'Hagan, J., Guerra-Gomez, J. A., Williamson, J. H., Cesar, P., and Shamma, D. A. (2022). "Digital proxemics: Designing social and collaborative interaction in virtual environments," in *CHI conference on human factors in computing systems* (New York, NY, USA: Association for Computing Machinery), 1–12. doi:10.1145/3491102.3517594

Yee, N., Bailenson, J. N., Urbanek, M., Chang, F., and Merget, D. (2007). The unbearable likeness of being digital: The persistence of nonverbal social norms in online virtual environments. *CyberPsychology Behav.* 10, 115–121. doi:10.1089/cpb.2006.9984

Yoshimura, A., and Borst, C. W. (2021). A study of class meetings in VR: Student experiences of attending lectures and of giving a project presentation. *Front. Virtual Real.* 2, 648619. doi:10.3389/frvir.2021.648619

Zibrek, K., and McDonnell, R. (2019). "Social presence and place illusion are affected by photorealism in embodied VR," in *Motion, interaction and games* (New York, NY, USA: Association for Computing Machinery). doi:10.1145/3359566.3360064

9

List of All Publications

2017

Michael Bonfert, Inga Lehne, Ralf Morawe, Melina Cahnbley, Gabriel Zachmann, and Johannes Schöning. 2017. **Augmented Invaders: A Mixed Reality Multiplayer Outdoor Game.** In Proceedings of the 2017 Symposium on Virtual Reality Software and Technology (VRST ’17), Association for Computing Machinery. DOI: 10.1145/3139131.3141208

2018

Mette Boldt, Michael Bonfert, Inga Lehne, Melina Cahnbley, Kim Korsching, Ioannis Bikas, Stefan Finke, Martin Hanci, Valentin Kraft, Boxuan Liu, Tram Nguyen, Alina Panova, Ramneek Singh, Alexander Steenbergen, Rainer Malaka, and Jan Smeddinck. 2018. **You Shall Not Pass: Non-Intrusive Feedback for Virtual Walls in VR Environments with Room-Scale Mapping.** In Proceedings of the 2018 Conference on Virtual Reality and 3D User Interfaces (IEEE VR ’18), IEEE Computer Society.
DOI: 10.1109/VR.2018.8446177

Michael Bonfert, Maximilian Spliethöver, Roman Arzaroli, Marvin Lange, Martin Hanci, and Robert Porzel. 2018. **If You Ask Nicely: A Digital Assistant Rebuking Impolite Voice Commands.** In Proceedings of the International Conference on Multimodal Interaction (ICMI ’18), Association for Computing Machinery. DOI: 10.1145/3242969.3242995

2019

Michael Bonfert, Robert Porzel, and Rainer Malaka. 2019. **Get a Grip! Introducing Variable Grip for Controller-Based VR Systems.** In Proceedings of the 2019 Conference on Virtual Reality and 3D User Interfaces (IEEE VR ’19), IEEE Computer Society.
DOI: 10.1109/VR.2019.8797824

2020

Dmitry Alexandrovsky, Susanne Putze, Michael Bonfert, Sebastian Höffner, Pitt Michelmann, Dirk Wenig, Rainer Malaka, and Jan David Smeddinck. 2020. **Examining Design Choices of Questionnaires in VR User Studies.** In Proceedings of the 2020 Conference on Human Factors in Computing Systems (CHI '20), Association for Computing Machinery. DOI: 10.1145/3313831.3376260

Nima Zargham, Michael Bonfert, Georg Volkmar, Robert Porzel, and Rainer Malaka. 2020. **Smells Like Team Spirit: Investigating the Player Experience with Multiple Interlocutors in a VR Game.** In Extended Abstracts of the 2020 Annual Symposium on Computer–Human Interaction in Play (CHI PLAY '20), Association for Computing Machinery. DOI: 10.1145/3383668.3419884

2021

Nima Zargham, Michael Bonfert, Robert Porzel, Tanja Doring, and Rainer Malaka. 2021. **Multi-Agent Voice Assistants: An Investigation of User Experience.** In Proceedings of the 2021 International Conference on Mobile and Ubiquitous Multimedia (MUM '21), Association for Computing Machinery. DOI: 10.1145/3490632.3490662

Michael Bonfert, Nima Zargham, Florian Saade, Robert Porzel, and Rainer Malaka. 2021. **An Evaluation of Visual Embodiment for Voice Assistants on Smart Displays.** In Proceedings of the 2021 Conference on Conversational User Interfaces (CUI '21), Association for Computing Machinery. DOI: 10.1145/3469595.3469611

Leon Reicherts, Nima Zargham, Michael Bonfert, Yvonne Rogers, and Rainer Malaka. 2021. **May I Interrupt? Diverging Opinions on Proactive Smart Speakers.** In Proceedings of the 2021 Conference on Conversational User Interfaces (CUI '21), Association for Computing Machinery. DOI: 10.1145/3469595.3469629

2022

Carolin Stellmacher, Michael Bonfert, Ernst Kruijff, and Johannes Schöning. 2022. **Triggermuscle: Exploring Weight Perception for Virtual Reality Through Adaptive Trigger Resistance in a Haptic VR Controller.** Frontiers in Virtual Reality 2.
DOI: 10.3389/frvir.2021.754511

Michael Bonfert, Stella Lemke, Robert Porzel, and Rainer Malaka. 2022. **Kicking in Virtual Reality: The Influence of Foot Visibility on the Shooting Experience and Accuracy.** In Proceedings of the 2022 Conference on Virtual Reality and 3D User Interfaces (IEEE VR '22), IEEE Computer Society. DOI: 10.1109/VR51125.2022.00092

Thomas Muender, Michael Bonfert, Anke Reinschluessel, Rainer Malaka, and Tanja Döring. 2022. **Haptic Fidelity Framework: Defining the Factors of Realistic Haptic Feedback for Virtual Reality.** In Proceedings of the 2022 Conference on Human Factors in Computing Systems (CHI '22), Association for Computing Machinery. DOI: 10.1145/3491102.3501953

Nima Zargham, Leon Reicherts, Michael Bonfert, Sarah Theres Voelkel, Johannes Schoening, Rainer Malaka, and Yvonne Rogers. 2022. **Understanding Circumstances for Desirable Proactive Behaviour of Voice Assistants: The Proactivity Dilemma.** In Proceedings of the 2022 Conference on Conversational User Interfaces (CUI '22), Association for Computing Machinery. DOI: 10.1145/3543829.3543834

2023

Michael Bonfert, Maiko Hübinger, and Rainer Malaka. 2023. **Challenges of Controlling the Rotation of Virtual Objects with Variable Grip Using Force-Feedback Gloves.** Frontiers in Virtual Reality 4. DOI: 10.3389/frvir.2023.1190426

Michael Bonfert, Anke V. Reinschluessel, Susanne Putze, Yenchin Lai, Dmitry Alexandrovsky, Rainer Malaka, and Tanja Döring. 2023. **"Seeing the faces is so important"—Experiences from online team meetings on commercial virtual reality platforms.** Frontiers in Virtual Reality 3. DOI: 10.3389/frvir.2022.945791

Malte Borgwardt, Jonas Boueke, María Fernanda Sanabria, Michael Bonfert, and Robert Porzel. 2023. **VRisbee: How Hand Visibility Impacts Throwing Accuracy and Experience in Virtual Reality.** In Extended Abstracts of the 2023 Conference on Human Factors in Computing Systems (CHI EA '23), Association for Computing Machinery. DOI: 10.1145/3544549.3585868

2024

Michael Bonfert, Thomas Muender, Ryan P. McMahan, Frank Steinicke, Doug Bowman, Rainer Malaka, and Tanja Döring. 2024. **The Interaction Fidelity Model: A Taxonomy to Distinguish the Aspects of Fidelity in Virtual Reality.** Preprint on arXiv. arXiv:2402.16665v1 [cs.HC]