Enhanced Virtuality: Increasing the Usability and Productivity of Virtual Environments

zur Erlangung des akademischen Grades eines

Doktors der Ingenieurwissenschaften

von der KIT-Fakultät für Informatik
des Karlsruher Instituts für Technologie (KIT)
(gemäß der Promotionsordnung vom 12. Januar 2017)

genehmigte

Dissertation

von

Adrian Heinrich Hoppe

aus Paderborn

Tag der mündlichen Prüfung:	10.05.2021
Erster Gutachter:	Prof. Dr.-Ing. Rainer Stiefelhagen
Zweiter Gutachter:	Prof. Dr. Bernd Fröhlich

Abstract

Current virtual reality (VR) hardware is on the verge of becoming a mainstream technology as display resolution and tracking accuracy increase steadily and prices fall. Different tools and frameworks help developers to create complex interactions with multiple users within adaptive virtual environments (VEs). However, as the landscape of VR grows, additional challenges arise: Diverse input controls with unfamiliar shapes and button layouts prohibit intuitive interaction. Furthermore, the limited functional range of existing software forces users to fall back to conventional personal computer (PC) or touch-based systems. Collaboration with other co-located users has challenges regarding tracking space calibration and collision avoidance. While working remotely, the interaction is influenced by latency and connection losses. Lastly, users have different requirements for the visualization of content, e.g., size, orientation, or appearance, within the virtual worlds. Strictly replicating real-world conditions within VEs will not allow to accommodate the individual needs of users.

To address these issues, this thesis presents advancements within the domains of input, collaboration, and enhancement of virtual worlds and users that target to increase the usability and productivity of VEs. First, PC-based hardware and software are transferred into the VE to foster the familiarity and functional range of existing applications. Virtual proxies of physical devices, i.e., keyboard and tablet, and a VR-mode for applications allow users to transfer real-world skills into the virtual world. Next, an algorithm is presented that allows the calibration of multiple co-located VR devices with high accuracy and low hardware requirements and effort. The pertinence of a full-body avatar visualization is validated for co-located collision avoidance and remote collaboration as VR headsets block out the real-world surroundings of users. Moreover, personalized spatial or temporal modifications of the VE are proposed that allow increasing the usability, task performance, and social presence of users. Discrepancies between the virtual worlds due to personal adaptations are compensated using avatar redirection methods. Finally, some of the methods and insights are integrated into an exemplary application to highlight their practical applicability.

The presented work shows that VEs can build upon real-world skills and experiences to enable a familiar and easy interaction and collaboration of users. Furthermore, personal enhancements of the virtual content and user avatars allow to overcome real-world limitations and boost VR experiences.

Kurzfassung

Mit stetig steigender Bildschirmauflösung, genauerem Tracking und fallenden Preisen stehen Virtual Reality (VR) Systeme kurz davor sich erfolgreich am Markt zu etablieren. Verschiedene Werkzeuge helfen Entwicklern bei der Erstellung komplexer Interaktionen mit mehreren Benutzern innerhalb adaptiver virtueller Umgebungen. Allerdings entstehen mit der Verbreitung der VR-Systeme auch zusätzliche Herausforderungen: Diverse Eingabegeräte mit ungewohnten Formen und Tastenlayouts verhindern eine intuitive Interaktion. Darüber hinaus zwingt der eingeschränkte Funktionsumfang bestehender Software die Nutzer dazu, auf herkömmliche PC- oder Touch-basierte Systeme zurückzugreifen. Außerdem birgt die Zusammenarbeit mit anderen Anwendern am gleichen Standort Herausforderungen hinsichtlich der Kalibrierung unterschiedlicher Trackingsysteme und der Kollisionsvermeidung. Beim entfernten Zusammenarbeiten wird die Interaktion durch Latenzzeiten und Verbindungsverluste zusätzlich beeinflusst. Schließlich haben die Benutzer unterschiedliche Anforderungen an die Visualisierung von Inhalten, z.B. Größe, Ausrichtung, Farbe oder Kontrast, innerhalb der virtuellen Welten. Eine strikte Nachbildung von realen Umgebungen in VR verschenkt Potential und wird es nicht ermöglichen, die individuellen Bedürfnisse der Benutzer zu berücksichtigen.

Um diese Probleme anzugehen, werden in der vorliegenden Arbeit Lösungen in den Bereichen Eingabe, Zusammenarbeit und Erweiterung von virtuellen Welten und Benutzern vorgestellt, die darauf abzielen, die Benutzerfreundlichkeit und Produktivität von VR zu erhöhen. Zunächst werden PC-basierte Hardware und Software in die virtuelle Welt übertragen, um die Vertrautheit und den Funktionsumfang bestehender Anwendungen in VR zu erhalten. Virtuelle Stellvertreter von physischen Geräten, z.B. Tastatur und Tablet, und ein

VR-Modus für Anwendungen ermöglichen es dem Benutzer reale Fähigkeiten in die virtuelle Welt zu übertragen. Des Weiteren wird ein Algorithmus vorgestellt, der die Kalibrierung mehrerer ko-lokaler VR-Geräte mit hoher Genauigkeit und geringen Hardwareanforderungen und geringem Aufwand ermöglicht. Da VR-Headsets die reale Umgebung der Benutzer ausblenden, wird die Relevanz einer Ganzkörper-Avatar-Visualisierung für die Kollisionsvermeidung und das entfernte Zusammenarbeiten nachgewiesen. Darüber hinaus werden personalisierte räumliche oder zeitliche Modifikationen vorgestellt, die es erlauben, die Benutzerfreundlichkeit, Arbeitsleistung und soziale Präsenz von Benutzern zu erhöhen. Diskrepanzen zwischen den virtuellen Welten, die durch persönliche Anpassungen entstehen, werden durch Methoden der Avatar-Umlenkung (engl. redirection) kompensiert. Abschließend werden einige der Methoden und Erkenntnisse in eine beispielhafte Anwendung integriert, um deren praktische Anwendbarkeit zu verdeutlichen.

Die vorliegende Arbeit zeigt, dass virtuelle Umgebungen auf realen Fähigkeiten und Erfahrungen aufbauen können, um eine vertraute und einfache Interaktion und Zusammenarbeit von Benutzern zu gewährleisten. Darüber hinaus ermöglichen individuelle Erweiterungen des virtuellen Inhalts und der Avatare Einschränkungen der realen Welt zu überwinden und das Erlebnis von VR-Umgebungen zu steigern.

Danksagung

Während meines Studiums am Karlsruher Institut für Technologie (KIT) entwickelte ich ein großes Interesse an der Gestaltung von Schnittstellen für die Mensch-Maschine-Interaktion. Durch die Vorlesungen von Dr. Jürgen Geisler kam ich zum ersten Mal mit dem Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung (IOSB) in Kontakt. Später begeisterte mich Dr. Florian van de Camp in einer Vorlesung für das Thema Virtual Reality, worauf ich mich um eine Masterarbeit am IOSB bewarb. Hieraus resultierte schließlich eine mehrjährige Forschungsarbeit, die zu dieser Dissertation führte.

Ich bedanke mich herzlich bei Prof. Dr. Rainer Stiefelhagen für die Möglichkeit der Promotion und für die Betreuung und Unterstützung während dieser Zeit. Des Weiteren danke ich Prof. Dr. Bernd Fröhlich für die Übernahme des Zweitgutachtens. Mein außerordentlicher Dank gilt Dr. Florian van de Camp, der mich während der gesamten Promotionszeit immer hervorragend unterstützt hat. Die konstruktiven und anregenden Diskussionen über die Themen der Dissertation sowie Abseits der Arbeit habe ich immer sehr geschätzt. Zudem danke ich Sebastian Maier für sein offenes Ohr und die Hilfe bei der Lösung komplizierter Herausforderungen und die immer angenehme Gesellschaft. Ich danke allen Kolleginnen und Kollegen des IOSB und des KIT für die kreativen und zielführenden Gespräche. Außerdem danke ich Prof. Dr. Jürgen Beyerer und Dr. Elisabeth Peinsipp-Byma für die finanzielle Unterstützung meiner Arbeiten im Rahmen mehrerer Forschungsprojekte. Meinen Eltern Ilona und Bernd, meinem Bruder Michael, meinen Großeltern Gerti und Heinrich, meinem Freund David und der Familie Rüthing danke ich für die Ermutigungen während der gesamten Zeit. Schließlich gilt mein ganz besonderer Dank meiner liebsten Christina, die mir stets Rückhalt und Kraft gibt. Du bist mein Alles und Segen, immerzu.

Contents

Appendix

1 Introduction

On December 9th, 1968, Douglas Engelbart presented the first prototype of the desktop computer mouse during the *mother of all demos* [Eng68]. Not five years later, on March 1st, 1973, the first mouse-based desktop computer with a graphical user interface (GUI), the Xerox Alto[1], was released. The mouse interaction combined with the windows, icons, menus and, pointer (WIMP) metaphor [Kim09] provides the foundation for the way that users interact with computer systems to this day. In October 1965, Eric Johnson [Joh65], presented his first idea for a touch-based input which would later be implemented into mobile devices including smartphones, e.g., IBM Simon in 1993[2], to become the arguably most ubiquitous digital device of current times.

Almost ten years earlier, on May 24th, 1957, Morton Heilig applied for a patent [Hei57] for the first virtual reality (VR) head-mounted display (HMD). The device included a stereoscopic video and audio display. Following the concept of the *ultimate display* [Sut65], Ivan Sutherland and Thomas Furness extended this idea and presented the *sword of Damocles* in 1968 [Sut68]. Their VR HMD included head tracking to adjust the displayed three-dimensional (3D) virtual environment (VE) to match the user's movements. Thomas DeFanti and Daniel Sandin then developed the Sayre Glove in 1976 [Stu94a] to transfer the dexterity and naturalness of hand interaction into VEs.

---

[1] https://www.parc.com/about-parc/parc-history/

[2] https://www.microsoft.com/buxtoncollection/detail.aspx?id=40

Although all these technologies have their origin at roughly the same time, past VR systems failed to establish themselves for widespread use beyond specialized fields of research, arcade gaming, and industry. However, the development of the Oculus Rift[3] by Palmer Luckey sparked a *second wave* of VR [Ant16]. Building on the increased hardware performance, higher display resolutions, and smaller form factors, new inventions such as cable-based or mobile HMDs, haptic devices, controllers, vests, omnidirectional treadmills, and tracking technologies jump-started the efforts of industry and research to establish VR at a consumer level. VR allows to visualize complex data [Hel14, Mar14a], create or validate designs [Rix16] and collaborate with remote experts [Gre17a, Lux16, Ngu16]. However, for many years, VR remained a specialized technology for specific use cases. No universal standard for menu structure or interaction was established to allow VR technology to be utilized for a variety of application areas. As no standardized method exists, there is lots of room to explore and optimize current systems. The following sections discuss challenges in the areas of input, multi-user collaboration, and modifications within the virtual worlds.

1.1 Input

VR allows users to interact with complex virtual worlds using magic wands/ controllers with buttons and joysticks, glove- or optical tracking-based gestures, or gaze controls using eye-tracking. None of these technologies has established a universal method of interaction. As no standards for interaction like the WIMP interaction for desktop computing exist, each application defines different controls that are custom-tailored to the respective use case. Users have to adapt to the input technology and control scheme each time. Furthermore, not all applications support each input device and HMD configuration which adds to the confusion.

---

[3] https://www.oculus.com/rift/

VR systems demand specially designed applications that support the stereoscopic rendering and gesture-, gaze-, or wand-based interaction. The required efforts to transfer an existing application into a VE are quite high. It is not enough to copy the source code into a VR-capable game engine as the existing presentation and control of the GUI have to be reworked. Furthermore, some applications are well suited for two-dimensional (2D) visualization and interaction, e.g., a web browser or spreadsheet application, and they might not even benefit from the 3D space. However, users might want to have access to 2D applications while being immersed in the virtual world to e.g., browse the internet for information or check their calendar. Current personal computer (PC)-powered VR headsets allow the user to access the desktop as a 2D window and interact with it using ray-casting interaction [Bow97b]. But, those approaches are cumbersome, not sufficiently integrated, and do not fully explore the possibilities of improving the visualization of and the interaction with 2D software.

1.2 Multi-User Collaboration

A collaborative virtual environment (CVE) allows multiple users to analyze and discuss information as well as interact with the VE and each other [Ben01]. VR is suitable for multi-user cooperation because of the natural visualization and interaction, e.g., [Fré04, Kol18, Le 16b, Oda15, Otm11, Ott06, Par00, Wan19b]. Users that engage in collaborative work using VR achieve better results [Hel06, Sch16c] especially when using full-body gestures [Dod10]. CVEs facilitate richer and more positive interaction and improve team cohesion [Ven12]. Additionally, VR provides more effective and efficient communication as teams can leverage affordances, e.g., visual cues, environmental feedback, and a sense of presence [Mon11]. Collaboration can either be co-located (on-site) or remote (in different rooms). The meta-review of collaborative mixed reality systems from 2013 − 2018 of de Belen et al. [Bel19] shows that 103 of 259 systems use a co-located collaboration of multiple users in a shared physical environment.

Working in the same room allows users to exchange tools and provide haptic feedback to each other. Co-located users can talk directly without any audio connection and only one area needs to be freed and equipped with tracking sensors. To enable co-located collaboration, e.g., tool exchange, the virtual worlds of the different users need to be calibrated. Systems with six degrees of freedom (DOF) employ tracking techniques using sensors such as inertial measurement units (IMUs), RGB or infrared cameras, and/or depth sensors. This allows each device to track its location and rotation in a self-defined coordinate system to enable accurate perspective changes depending on the user's head movement. However, there exists no standardized way to calibrate different VR and augmented reality (AR) systems as current services are only available for specific devices and not extendable to other platforms.
Another disadvantage of the co-located approach is the possibility of unintended user contact, which can damage the hardware or even hurt the involved parties. Sufficient user representation is needed to enable users to successfully predict the other person's position to avoid collisions.

Based on the confirmed advantages of a 3D CVE and the increasing interconnection of our world, it makes sense to utilize CVEs also remotely. Remote collaboration allows teams to spontaneously discuss a topic, or experts/teachers to support workers/students with problems during task execution. Furthermore, this saves time and resources while protecting the environment and climate as people can communicate quickly and independently of their physical location [Gre17b]. Even though VR enables co-located and remote collaboration, the impact of the user's physical location has yet to be determined. One disadvantage of a remote interaction can be a delayed interaction due to network latencies or even connection loss. However, the immersive visualization of VR may provide new ways to compensate for network issues, e.g., computer-controlling an absent player to mask her/his disconnect.

1.3 Enhanced Collaboration

VR offers the possibility to experience adventures that are exhausting, rare, or dangerous [Gre17b], all while sitting comfortably in a chair. Without ever leaving the house, one can climb Mount Everest[4] or visit space[5]. But, if the virtual world is only based on the laws of physics and tries to emulate reality, it wastes potential. VR can enhance reality by rolling back time or defying the laws of physics all at once. Slater and Sanchez-Vives [Sla16] put it like this: “Throughout and wherever possible we have stressed [...] how the real power of VR is not necessarily to produce a faithful reproduction of ‘reality’ but rather that it offers the possibility to step outside of the normal bounds of reality and realize goals in a totally new and unexpected way”.

Besides controlling matter and time, VE can also be adjusted to display individual content for each user and optimize the human-to-human interaction (HHI). A natural way to convey intent is to use voice communication together with gaze, gestures, facial expressions, or full-body movements. While working in a face-to-face configuration, users can see the virtual avatar of other users and the virtual content at the same time. However, when users work with different perspectives, their task performance can decrease due to left-right ambiguities [Fei18]. In contrast, when standing side-by-side or even at the same location, users can use the shared perspective to build a common understanding. Users see objects from the same angle and can be sure that the other person sees the same. But, this might cause the avatars of the different users to overlap, which might be awkward. One way to combine the advantages of both, face-to-face cooperation and a shared perspective, is to switch between the two configurations. But, the switch takes time and users may want to be in different configurations simultaneously. Standing face-to-face and at the same location is impossible in the real world.

Virtual worlds allow full control over the experience each user has. VR technology can enhance reality to make it more understandable and more accessible. It allows users to share an environment that is totally adapted to their very needs. Reducing the complexity of a task by rotating content is one possible adaption. Furthermore, visual impairment may be addressed with

---

[4] http://www.solfar.com/everest-vr/

[5] https://505games.com/games/adr1ft/

up-scaled content that contains bigger fonts and adjusted contrast. The effect of deaf-muteness might be reduced with text-to-speech and speech-to-text functionality. The challenge is to combine all the different views and personal requirements into one consistent environment to enable a truly enhanced collaboration between users.

1.4 Outline, Goals, and Contributions

The goal of this thesis is to increase the usability and productivity of VEs. To achieve this target, several concepts are explored.

However, first of all, Chapter 2 provides an overview of existing related work in the domains of perception, input, and collaboration of VR systems. It sets the stage for this work and explains the reasoning behind the topics presented. Additionally, Appendix A describes the statistical methods and questionnaires which are used for the evaluation of the presented systems.

Chapter 3 presents methods that aim to increase the input performance and scope of VR applications. Easy-to-learn, efficient and reliable input with a variety of applications is one of the key factors that limit the widespread adoption of VR. With the use of virtual proxies, real-world physical devices are transferred into the virtual world to provide visual and haptic feedback for users. As these proxies represent commonly used input devices, users can transfer their real-world skills into the VE. Current VR software is specially designed and implemented for specific hardware. Therefore, the available functionality of VEs is limited when compared to PC or smartphone systems. To increase the functional range of VR, a mechanism to transfer arbitrary 2D desktop applications into VR is provided. The proposed method allows integrating existing applications as-is into the virtual worlds. Further extensions allow transferring suitable parts of the 2D applications little by little towards a 3D interface while retaining a fully functioning application at all times.

Chapter 4 presents several mechanisms to enable multi-user collaboration in co-located and remote setups. Collaborative work has many advantages. However, the requirements to enable collaboration are high. Two or more co-located users cannot interact directly with each other if the coordinate systems of their tracking devices are not calibrated. Current VR HMDs do not offer sufficient support to align the different virtual worlds of users as several sensor/tracking technologies and operating systems are used. Therefore, an algorithm is presented that allows calibrating two or more arbitrary 6DOF devices toward a shared coordinate system. As users start to explore the VE collaboratively, they need to be aware of the other user's position to avoid collisions. The presented work validates the importance of a full-body avatar for collision avoidance. In addition to that, the full-body avatar in combination with speech communication can also enable remote collaboration. A comparison of a co-located and remote task execution shows that the immersion of VR is so strong that it allows users to collaborate independently of their physical presence.

Chapter 5 explores spatial and temporal manipulations of CVEs to increase usability, task performance, feeling of teamwork, and perceived co-presence to yield a new way of easy and efficient teamwork with a high social presence. For this, a mechanism is presented that allows to independently adapt the virtual worlds to the needs of the respective users. Synchronization of content and avatar adaptation strategies maintain the interrelation between the different virtual worlds. These manipulations build upon redirection techniques and enable multiple users with full-body avatars to cooperate in 3D tasks that involve object selection and manipulation. Furthermore, temporal modifications allow compensating short-term interruptions of a conversation by masking the intermittence and replaying the missed speech and avatar movements. This allows users to rejoin a conversation seamlessly.

Chapter 6 presents an exemplary application that integrates many of the systems or insights gained within previous chapters. It highlights the practical applicability of the proposed methods.

Finally, Chapter 7 concludes this work and provides an outlook beyond the scope of this thesis.

The main contributions of this work are:

- Design, implementation, and evaluation of two virtual proxy input modalities that provide haptic feedback on movable devices with a moving user, without wearable trackers, and with commodity hardware to enable efficient input based on skill-transfer from the real world into the VE [Hop18a, Hop19, Hop20b] (see section 3.1).
- A hybrid interaction system that increases the functional range of VR by enabling the interaction with existing arbitrary 2D applications that are partially transferred to and uniformly controlled within VEs using a VR-mode [Hop20c] (see section 3.2).
- Design, implementation, and evaluation of a novel algorithm for precise, easy-to-learn, easy-to-perform, and low-power calibration of two arbitrary 6DOF devices that enables co-located collaboration [Hop20a] (see section 4.1).
- The validation of a full-body avatar for avoiding collisions of co-located users and enabling collaboration independent of their physical location [Hop18b, Hop18c] (see sections 4.2 and 4.3).
- Design, implementation, and evaluation of a social redirection technique that enables multi-user collaboration with full-body avatars while providing each user with an individual optimal perspective to increase task performance and social presence in comparison to unmodified or baseline methods [Hop18d, Hop21] (see chapter 5).

2 Related Work

There exist different definitions of what VR is. Heim [Hei00] defines it based on the three factors of immersion, interaction, and information intensity. *Immersion* describes the feeling of being transported to another world. It is achieved by overwriting/misleading the user's senses. *Interaction* describes the capability of changing the virtual world based on the user's input [Nal06], e.g., navigation within the VE or manipulation of objects. Interaction is influenced by factors such as speed, range, and mapping [Ste92]. The quicker a system reacts to the user's input the better. Range describes the number of interactable objects. Mapping demands that the actions of a user within the real-world should be matched with appropriate actions in the VE. Lastly, *information intensity* describes the knowledge transfer from the virtual world to the user [Nal06].

Steuer [Ste92] finds that VR can be described without technical definitions but via the experience of presence and telepresence. He defines *presence* as the experience and perception of a physical environment via the human senses and mental processes. *Telepresence* describes the experience of presence within a mediated environment. This mediated world can differ from the user's real world in space and time. VR then can be defined as an environment in which the user experiences telepresence. Oh et al. [Oh18] extend this definition towards *social presence*. Social presence defines "the feeling of being there with a 'real' person" and has a positive impact on communication.

Other definitions include visual, acoustic, and haptic displays which immerse the user using a rendering engine and tracking systems [Bro99]. Green and Jacob [Gre91] emphasize the real-time requirements of the diverse input and output devices of VR systems.
Users can train workflows in safe and controlled VEs [Muj04]. VR finds use

in the areas of data visualization, modeling, designing and planning, training and education, cooperative working, psychiatric treatment, tourism, and entertainment [Azu97, Bow04, Haa06, Maz96].
Most VEs are implemented using a game engine such as Unity[1] or Unreal Engine 4[2]. However, depending on the hardware and software requirements, specialized systems are also commonly used.

To further clarify the characteristics and components of VR, this chapter presents related work regarding perception and input. As multi-user collaboration is an important factor of productive work, this topic is also addressed. Different collaborative systems are presented, as well as key factors to increase work productivity, and adaptations of the virtual worlds to further increase productivity. This chapter serves as a foundation for the following chapters and builds upon [Hop18a, Hop18b, Hop18c, Hop18d, Hop19, Hop20a, Hop20b, Hop20c, Hop21, van20].

2.1 Perception of Virtual Environments

Bowman et al. [Bow01d] define four classes of displays for VR systems. *Optical full-immersive* displays, such as HMDs, arm-mounted displays, or virtual retinal displays fully mask the real world. *Optical semi-immersive* displays, such as stereo-monitors, workbenches, surround screens, or cave automatic virtual environments (CAVEs) [Bro99] allow users to see the real and virtual world in parallel, e.g., their own body within a VE. *Auditory* displays provide 3D sound and allow users to localize the spatial position of an audio source. Finally, *tactile* or *haptic* displays provide force-feedback or touch-feedback [Bur96, Shi93] and let users experience effects like gravitation, damping, friction, or the texture of surfaces.
As VR display systems overwrite the real world and emulate a different world, the depth perception of users is changed [Mar19]. This leads to depth underestimation for objects in the space of action (2 – 30 m) and distant space (>30 m)

---

[1] https://unity.com/

[2] https://www.unrealengine.com/

as well as depth overestimation for objects in personal space (<2 m). However, seeing a fully-articulated and tracked visual representation of oneself leads to more accurate depth estimations within 4 to 6 meters [Moh10]. Also, working within arms-reach adds additional information through head motion parallax and proprioception [Min97a].

Current HMD devices are cheap, small, and mobile. The integrated lenses warp the displayed content to provide the user with a stereoscopic impression of the virtual world. Basic systems, such as a Google Cardboard[3] include 3DOF tracking, i.e., rotation, and allow users to look around the VE. 3DOF tracking can be achieved using orientation sensors, e.g., an IMU. More advanced systems, such as an HTC Vive (Pro)[4], Oculus Rift (S)[5], or Valve Index[6] employ 6DOF tracking, i.e., rotation and position, which allows users to experience the virtual world by moving their head or even walking around. 6DOF tracking systems use optical, electromagnetic, or acoustic sensors [Che16a, Par19, Spi14, Stu94b]. Tracking can be achieved using an inside-out or outside-in tracking method. *Inside-out* tracking employs sensors on the device to recognize its position and orientation, e.g., a camera using depth information [Ha14] or simultaneous localization and mapping (SLAM) [Dao18], or multiple strategies [Bra15]. *Outside-in* tracking requires hardware to be set-up in the physical environment of the user. This hardware either captures the scene to find the device's pose, e.g., Microsoft Kinect[7] or Vicon[8], or sends out signals which can be interpreted by the devices to determine their position and rotation, e.g., Lighthouse[9].
VR tracking and output need to present the virtual world very accurately and with the lowest latency possible [All01]. Otherwise, users may experience motion sickness which may lead to nausea. Motion sickness occurs when

---

[3] https://arvr.google.com/cardboard/

[4] https://www.vive.com/

[5] https://www.oculus.com/

[6] https://store.steampowered.com/valveindex/

[7] https://developer.microsoft.com/windows/kinect/

[8] https://www.vicon.com/

[9] https://partner.steamgames.com/vrlicensing#Tracking

users are presented with conflicting information of the visual and vestibular senses [Aki03].

The VE allows objects to be spatially distributed to support the meaning and relationship of information. Users tend to place virtual windows on walls, grouped by their content, frequency of use, and context (e.g., work or home) [Rob00]. Especially women benefit from the higher field of view (FOV) of larger displays [Cze02].
Feiner et al. [Fei93] and Lindeman et al. [Lin99] differentiate between three categories of window positioning. *Surround-fixed* or *world-fixed* windows have an absolute position in the room and do not move. *Display-fixed* or *view-fixed* windows are fixed to the user's view and move with the head. Lastly, *world-fixed*[10] or *object-fixed* windows assume a fixed position relative to another object in the VE, e.g., a cube or the hand of the user. Bowman et al. [Bow08] describe that objects should mostly be attached to other objects. Additionally, it should be avoided that virtual objects intersect with each other as this is not possible in the real world.
LaViola and Keefe [LaV11] distinguish between three types of windows. *Hand-oriented* menus are attached to the user's hand. *Converted* 2D menus are based on conventional desktop windows and the WIMP metaphor. Finally, *3D widgets* integrate their functionality into the VE using 3D objects. Widgets should utilize affordance (i.e., provide an intrinsic understanding of what is offered [Gib79]), be visually simple, and match the dimension of the controlled value (e.g., a slider for a scalar value) [Her94].

While working in the virtual world, users can be presented by a virtual avatar that performs movements according to their real-world motions. To enable a natural conversation, the user can be tracked and any facial or body expressions transferred into the VE [Fec16, Sey17]. With the gathered information, an avatar for each user of the VE can be rendered. Users experience a sense of embodiment while using virtual avatars. Kilteni et al. [Kil12a] define three factors of embodiment, i.e., the sense of self-location, the sense of agency, and

---

[10] Note the difference between the two interpretations of *world-fixed* by Feiner et al. and Lindeman et al.

the sense of body ownership. The *sense of self-location* describes the experience of being inside a body. *Agency* defines the sense of having intentional control over one's body. Consequently, it is reduced if the virtual avatar's movement does not match the real-world movements [Pad16]. The *sense of body ownership* describes the feeling that a body belongs to oneself. Its effect can be observed in the real [Bot98] and virtual [Gon14, San10, Yua10] world. A virtual body can significantly increase the user's presence, e.g., during locomotion [Sla95, Uso99]. Also, an avatar can increase the accuracy of height [Lin13] and distance [Moh08, Moh10, Rie08] estimations if calibrated to the correct size. Differently scaled avatars, e.g., child-size, lead to overestimation of size and altitude [Ban13]. Steed et al. [Ste16] show that having an active avatar alleviates the mental load of doing spatial rotation exercises and improves performance in a letter recall task. Virtual avatars enable collaboration that is similar to real-world face-to-face communication [Smi18]. Furthermore, the social behavior of users is changed, e.g., if provided with a more attractive avatar [Yee07a] or different clothing [Peñ09]. If the behavior of the virtual avatar does not match the user's real actions, users can have problems identifying themselves with their virtual self [Bio14]. However, humans seem to be flexible when adapting to a virtual body. Also, even though virtual avatars are beneficial, users perform quite well without them [McM11, Str09]. Kilteni et al. [Kil15] discuss and survey the extensive work that was performed in the domain of body ownership illusions.

2.2 Input in Virtual Environments

Bowman et al. [Bow01d] differentiate between discrete (e.g., button), continuous (e.g., tracking), and hybrid input. Current work can be grouped into device-based, gesture-based, or multimodal input techniques [Lep17]. *Device-based* techniques use a magic wand/controller [Cig03, Wlo95b] also with dynamic shapes [Shi19], gamepad [Go08, Iso04, Köl07, Per98, San06, Wil06, Wob04], phone [Gon09, Kim17b], keyboard [Bow02, Gon09, Kim04, Lin17, Man98, Wal17], pen and tablet [Bow02, Gon09, Pou98b] or touch [Che14b, Gro15, Lee16a]. *Gesture-based* methods use hand [Baj12, Bol19, Bow01b, Bow02, Eva99, Gon09, Hes11, Ni11, Ros99] and head gestures [Yu17]. *Multimodal* techniques often add speech [Bow02, Hos12, Mul98] or gaze [Hil16, Pfe15, Pfe16, Pfe17] to the physical devices.

VR input has multiple challenges. First, the precise manipulation of objects in-air is difficult if a system fails to provide sufficient haptic feedback [Min97a]. The lack of physical work surfaces for alignment and support limits users' precision and leads to fatigue. Second, a uniform interaction scheme containing input elements such as buttons or sliders is missing [Lar03, Min97a]. Third, Bowman et al. [Bow08] hypothesize that widespread adoption of VR is limited by the significant training effort that is required before the systems can be used productively.

2.2.1 Increasing Input Performance

Several techniques can be introduced to increase the usability of VR user interfaces (UIs). The methods presented below include the use of proprioception, nonlinear mappings, clutching, constraints, and haptic feedback.

Proprioception describes the sense of joint position/movement and forces that act upon them [Bof94]. It can be used as additional feedback within VEs to provide intuitive, precise, and comfortable interaction [Min97a]. Additionally, body-relative interactions can be performed without looking. Proprioception can be experienced using different methods, e.g., direct manipulation, physical mnemonics, or gestures. *Direct manipulation* allows to control objects by grabbing them, e.g., go-go [Bow99] or scaled world grab [Min97a]. *Physical mnemonics* allow users to utilize virtual objects, tools, or menus relative to their own bodies. The body serves as a memory aid, which allows users to recall where objects are stored and how they can be accessed, e.g., handheld windows [Lin99]/widgets [Min97a] or pull-down menus [Min97a]. Lastly, users can execute commands using hand or body *gestures*, e.g., head-butt zoom or look-at menus [Min96, Min97a]. Devices that enable finger manipulation instead of wrist, elbow, or shoulder movement increase users' performance [Min97a, Zha96]. One should note that users often perform tasks bimanual and each hand takes on a different role [Gui87]. Users initially specify a frame of reference using their non-dominant hand. Then, the dominant hand executes precise actions with respect to this reference frame.

A *nonlinear mapping* between the real and virtual movements of users helps to increase the precision or speed and reduce the fatigue of the interaction [Min96]. An overproportional movement allows quick actions or can increase the user's reach within the VE. Underproportional movements allow users to perform small and precise movements.
Working with arm gestures can be tiring. *Clutching* allows detaching the movement of the hand from the virtual movements completely. It allows users to pause the interaction and relax their body [Sto95, Wlo95a].

Virtual and physical *constraints* help to increase the precision of a user's interactions, especially in VR [Min95]. *Virtual constraints* reduce the dimensionality of the user's input by ignoring certain dimensions of movement, e.g., only allowing horizontal movement. However, movement in the ignored dimension is possible. In addition to that, snapping methods can further reduce the freedom of movement to discretized steps. *Physical constraints* limit the movement of the user by restricting physical movement, e.g., by introducing haptic surfaces or other methods to constrain the user's movement. But, these may reduce the user's capability of expression.

Haptic feedback increases user's performance and usability [Hin94, Kos08, Mar12, Oh20b, Pla20, Sal09, Tza08]. Devices that emit haptic feedback can enable interaction with diverse applications, e.g., sketching [Aro18, Dre20, Gas19, Huo17], exploration and annotation for 3D visualizations [Büs19, Lóp16, Son11], 3D input [Bab18, Dor16, Sur19], single-user object manipulation [Kat15, Lia13, Mar14b, Mos13, Spi10, Sur19, Vin16], and collaborative object manipulation [Gra17, Sal09]. Haptic feedback can be active, e.g., devices that actively exert pressure/forces on the user, or passive, e.g., devices that limit a user's movement. The active or passive devices can be tracked to integrate themselves into the VE. Users can experience haptic feedback using a variety of different methods. These include the usage of movable [Tak15, Tak18], body-mounted [Gug16, Lee16b], or surface-mounted [Sin13] virtual or real (touch) displays. Furthermore, touch can be enabled on a variety of surfaces, e.g., a ring [Ens16], a globe [Eng19], flexible surfaces of variable size [Hol14, Nit18, Pou19], the user's skin [Dez12, Har11, Lap14, Wei15, Xia18c, Zha19] or clothes [Dob15, Par17, Pes12]. Also, drones [Gom16, Hop18e] can act as a surface for touch feedback. Moreover, haptic actuators/exoskeletons attached to the user's hand [Kim16, Kim17a, Kon03, Sar06, Sch07, Sch15, Sol10, Tse14], as separate controllers [Chi12, Cho18, Kam09, Kyu09, Whi18], or as robot-arm devices [Kyu09, Mas94] can exert forces on the user. Araujo et al. [Ara16] use a robotic arm to move a haptic surface to block the user's hand when it gets close to a virtual object. The head of the robotic arm can be rotated and exchanged, hereby providing a wide range of haptic surfaces, i.e., different textures, a pressure sensor, an interactive surface, physical controls, or even a heat emitter. However, interacting with real touch surfaces, especially tablets, has the disadvantages of a limited surface size and that both hands are needed for interaction [Lar03]. Using virtual surfaces can have advantages as they can be transformed without any constraints [Loo07]. Costes et al. [Cos20] and Wang et al. [Wan19a] provide a good overview of haptic feedback systems.

Devices whose virtual representation, e.g., shape or weight, corresponds to their physical shape increase presence and tactile sensations [Wlo95a]. Simeone et al. [Sim15] show that the virtually displayed object does not need to match the real object perfectly. Some variations of material properties or shape allow the use of a broader range of virtual objects even if the set of real-world objects is limited. Corsten et al. [Cor13] also show that real-world objects can be repurposed for system control input, e.g., bottle cap as rotary nob or clicky pen as presenter button. Furthermore, transparent tablets can serve as magic lenses to allow magnification or filtering of content [Bro06, Lei15, Spi09, Tom14, Tom17] or to supply users with additional information [Gru15]. Larger magic lenses are more helpful than smaller ones [Oh06]. However, their aspect ratio is not so important.

2.2.2 Using Virtual Proxy Devices

Several systems use virtual proxy objects to provide haptic feedback. A virtual proxy is a digital replica that is superimposed over a real-world object [Rus97].

Virtual proxies can be used, for example, to recreate parts of a car's interior [Sal08], or to support storytelling [Har17]. Most important for this work are virtual replications of common input devices, i.e., keyboard and touch interfaces.

Although there are plenty of methods that achieve text input in VR, no technique has established itself as the defining standard. An analysis shows that the fastest techniques use a QWERTY keyboard (see Fig. 3.2). Error rates are lower when users type slower and concentrate on inputting single characters or when word correction is used (see Fig. 3.3). However, there are different ways to integrate a keyboard into the VE such as in-air gestures with [Wu17] or without [Ren13] gloves, ray-casting selection on virtual keyboards [Spe18], or physical keyboards [Bie20, Sch19]. A physical keyboard can use external [Gru18a, Gru18b, Kni18] or internal [Hop18a, Ott19] finger tracking. Furthermore, capacitive touch devices [Kim19b], pen and tablet devices [Spe18], or dull surfaces [Wan14] can be used for text input. While working in VR, users cannot see their hands on the keyboard. Different approaches are possible, where no hands, only the fingertips, a 3D virtual mesh representation of the hands, or an AR video see-through display are integrated into the VR view [Gru18a, Hop18a].

The arguably most commonly used interface for digital devices is the touch-based interaction. There exist several techniques that allow to present the user with an interactable surface. *Real screens* can be used to display content [Lei15, Lóp16, Mar14b, Mos13], for example in semi-immersive CAVE environments [Med13]. *Projection-based* systems can display content on dull surfaces. The projector system can either be static [Fuj19, Har12, Xia13, Xia16] or mobile [Kem18]. Also, the surface that is used for projection can be static, e.g., the top [Fuj19] or edge [Jos19] of a table, or mobile, e.g., movable handheld flat [Cha12, Spi09] or flexible [Ste13] surface, or the user's body [Har11]. Besides projecting on opaque surfaces, *transparent* objects can be used to allow users to interact with content that is actually displayed behind the surface, e.g., using a workbench display [Enc99, Sch02, Sch99]. *See-through* displays even allow creating 3D desktop experiences [Lee13]. Lastly, another approach is to use VR or AR technology to display a *virtual* screen on a haptic surface [Bow01a, Bow98, Che04, Lin99, Sza97, Xia18b, Zie19].

To enable interaction with haptic surfaces, touch positions need to be recognized. For this, the device itself can detect a touch, e.g., with a touch foil [Hol14, Wei15] that can be customized and printed [Gro18, Hol14, Sav12]. Capacitive touch detection can enable interaction on displays [Med13] or dull surfaces such as tables [Win17]. However, not all surfaces are augmented with touch detection. To detect a touch on these surfaces, tracking of the user's hands and the surface are necessary. The easiest solution when using static surfaces is to perform an initial calibration of the surface's position [Abd19]. However, to allow the surface to be moved freely, it needs to be tracked. External tracking can be enabled using Lighthouse tracking [Eng19, Sur19, Zha20], marker-based solutions [Med13], magnetic tracking [Lin99, Sch99, Ull97], or scene reconstruction [Iza11]. Surfaces can also be fixed to a movable actuator [Sin13, Tak15, Tak18]. Furthermore, internal sensors of the devices can be used to detect the device's pose, e.g., using a camera [Lei15, Lóp16, Mar14b, Mos13]. The user's hands can be tracked either by using a glove-based [Eng19, Lin99] or camera-based tracking method. Camera-based tracking can use single 2D cameras with structured light projections on a calibrated surface [Dai12, Dai14], distortion detection of a projected interface [Hu14], deep learning [Fuj19], or shadow monitoring of a hand on flat [Mat17] and nonplanar [Nii16] surfaces. Alternatively, one can use static depth cameras [Har12, Wil07, Wil10a], e.g., using a time of flight sensor [Par16, Zha16a] or low-cost hardware [Won16]. Also, thermal reflections can be used to detect in-air gestures [Sah14]. Advanced detection algorithms allow distinguishing multiple fingers, postures, users, and handedness in multi-user environments [Mur12]. Also, multiple cameras can be combined to increase the volume of the interaction [Wil10b] or to increase tracking accuracy using different types of sensors [Cad16]. The touch detection itself can occur using collision estimation through hand and surface tracking [Xia18b]. Dippon and Klinker [Dip11] found that camera-based touch detection is worse than capacitive touch detection but its accuracy is sufficient for interaction. Different camera angles can be used to increase touch detection accuracy, e.g., lateral camera position (side view) [Nte17, Qin16]. Touch can also be detected via the different grip poses that correspond to shape primitives [Zho20]. Additionally, camera and computer mouse-sensors at the fingertip [Yan12], wrist [Prä14, Yan20], or forearm [Sri17] can be used. The finger position and touch on the user's skin can also be detected with high frequency alternating current signals [Zha16b] or radio-frequency waveguides [Zha19]. Swept frequency capacitive sensing can enable touch on all conductive materials [Sat12], e.g., human body or liquids. Furthermore, an IMU can enable touch interaction through the detection of vibrations [Oh20a, Shi20]. Microphones on a wristband can also detect touch [Gon20]. Thermal cameras can detect touch by detecting the residual heat [Kur14] on surfaces. The touch feedback can further be enhanced using electrovibration haptic feedback [Zha20]. Finally, proactive touch estimation can be used to decrease touch latency [Xia14].

2.2.3 Interacting with 2D Interfaces

Bowman et al. [Bow04] describe that 2D tasks are cognitively easier to accomplish than 3D tasks. Therefore, using 2D windows and reducing the DOF of the interaction in VR might be a way to increase precision and usability and to decrease fatigue. Additionally, Bowman et al. [Bow08] note that interaction should only be allowed on visible objects. However, this does not mean that all tasks should be converted to 2D space. Darken and Durost [Dar05] find that interaction should match the dimensionality of the task. Tasks such as text reading or input should be carried out with 2D interfaces. Tasks that benefit from the third dimension, e.g., object selection or manipulation, should use 3D interaction. 2D interfaces facilitate ease of learning, speed, overview, and comprehension [Hep19]. An interaction in 3D is more fun and also facilitates comprehension. Hybrid interaction can allow users to switch between different systems depending on the current task. Hybrid interfaces may utilize a mouse and keyboard [Man18], gloves [Ben05], PDAs [Wat99], or tablets [Tre06, Wan15a, Wan15b] in combination with an immersive visualization. Also, virtual windows can provide common system controls [Di 03, Kec03, Ngu17] that can, for example, be controlled using hand gestures [Jac92] or virtual touch [Ang95a, Ang95b]. De Haan et al. [Haa06] demonstrate that

UI elements on 2D windows can be extruded to become 3D widgets. Regenbrecht et al. [Reg01, Reg02a, Reg02b] show that 2D input devices can be modified to enable 3D interaction, i.e., a mouse can be picked up to become a magic wand using marker-based tracking. Dachselt and Hübner [Dac07] survey various systems that include 2D or 3D menus within virtual worlds. They define different taxonomies of menus depending on properties such as their structure, purpose, placement, or content.

Several systems within the related work, e.g., [Con97, Toy18], or commercial applications use a ray-casting technique to interact with 2D virtual windows. However, ray-casting might not be the ideal solution because 3D interaction reduces the performance of traditional interaction techniques [Bow01d]. Ray-casting techniques can be difficult to use on smaller surfaces or larger distances because of their sensible reaction to changes in the rotation of the virtual laser pointer's origin. Techniques that try to minimize this precision problem are for example a bending ray-cast that always selects the closest interactable object [Ahl06, Ste05, van13]. However, when capturing an arbitrary application the positions and extends of buttons and other menu items might be unknown. Further modifications of the ray-cast try to increase the user's precision or decrease fatigue by using underproportional movements [And07b] or adjusting the position of the interactable surface to be more comfortable [And07a]. Other approaches position virtual windows, e.g., as a tablet, relative to the user's body [And07a, Ang95b, Lin99, Min97b, Wob09] or the environment [Dem99, Lar03] to utilize smaller distances or proprioception to increase precision.

To display a 2D application inside a VE without reimplementing the software for 3D, a screen capturing method can be used. The application is filmed, converted to a texture, and then displayed on top of an object, e.g., a (curved) plane, inside the VE. Microsoft Windows graphics device interface (GDI)[11] is an application programming interface, which captures the desktop or even single windows, and is used for example by the Open Broadcaster

---

[11] https://docs.microsoft.com/en-us/windows/win32/gdi/windows-gdi/

Software (OBS)[12]. Other systems include the Desktop Duplication API[13] or WindowsGraphicsCapture[14]. Related work shows that 2D windows can be integrated into the VE by running a full X [Fei93] or VNC [Bue03] server. Also, UI frameworks or applications such as Qt [And06] or a web browser [Bar05, Toy18] can be integrated into the virtual world. VR implementations may provide a user with a way to interact with the whole desktop or separate windows [Ull09]. Examples of current software for PC-powered VR HMDs that enable interaction with the user's desktop are SteamVR[15], Oculus Dash[16], Windows Mixed Reality home[17], Bigscreen[18], uDesktopDuplication[19], and Virtual Desktop[20].

2.3 Systems for Mixed Reality Collaboration

Immersive CVEs allow users to gather information about particular content or situations. Moreover, they enable users to discuss and analyze the gathered information together. Greenwald et al. [Gre17b] recap several advantages of VR for education/training: Access to remote experts, to scarce or access-limited resources, or to physically impossible experiences. They classify the actions in CVEs into interaction with other humans, and interaction with the environment. Slater and Sanchez-Vives [Sla16] discuss the advantages and disadvantages of VR collaboration in great detail. They give an overview of existing research on the domains of presence, science, education, training, remote collaboration, and shared environments. Proposed frameworks like MASSIVE [Ben97, Gre00], Studierstube [Sza98], PIT [Art98],

---

[12] https://obsproject.com/

[13] https://docs.microsoft.com/en-us/windows/win32/direct3ddxgi/desktop-dup-api/

[14] https://docs.microsoft.com/en-us/uwp/api/windows.graphics.capture/

[15] https://store.steampowered.com/app/250820/SteamVR/

[16] https://developer.oculus.com/documentation/native/pc/dg-dash/

[17] https://docs.microsoft.com/en-us/windows/mixed-reality/navigating-the-windows-mixed-reality-home/

[18] https://www.bigscreenvr.com/

[19] https://github.com/hecomi/uDesktopDuplication/

[20] https://www.vrdesktop.net/

and DIVE [Fré98] allow collaboration with co-located and distant users. For education, not only learning material is relevant but also the HHI since learning strongly depends on the exchange with others. Heldal et al. [Hel06] show that two people working as a team outperform individuals. Schouten et al. [Sch16c] verify that an avatar-based interaction in a 3D CVE supports team collaboration and decision making. The virtual world offers generally the same capabilities as text-based or audiovisual communication but allows for further functionality and yields a higher shared understanding and thus higher task performance in terms of consensus, satisfaction, and cohesion. Paul and Reddy [Pau10] show that collaborative understanding is negatively affected by ambiguous information and also by different roles and expertise of group members.

Collaboration can occur co-located, i.e., in the same room/on-site, and also remotely, i.e., at different locations. Hybrid configurations are also possible as several co-located users collaborate with other groups that are connected remotely [Bec13]. Walker et al. [Wal09] show that the performance and overall workload of users are similar when playing chess remotely or co-located. However, considerably different processes of team adaption are required when using mediated systems. Tang et al. [Tan11] found no differences in co-located pen-and-paper versus remote digital sketching collaboration. Sykownik et al. [Syk20] also report similarities between remote VR and unmediated (non-VR) co-located gaming based on a preliminary study. Born et al. [Bor19] argue that real-world co-located collaboration is better than mediated collaboration. However, VR systems allow to replicate users, objects, and environments and may enable a remote collaboration that is a suitable alternative to co-located work.

Several researchers explored the differences of the physical location of two users that collaborate within VR. Gómez Maureira and Verbeek [Góm16] found no significant differences between co-located or remote collaboration when using 3DOF HMDs regarding social presence, self-enjoyment, or immersion. Other work with 6DOF HMDs, also found no significant differences regarding presence, perceived co-presence, communication quantity, or cooperative interactions of co-located and distributed players [Pod18a, Sou20]. However, co-located user collision can significantly increase co-presence [Pod18a]. But, remote work can have significantly higher qualitative communication scores and partially higher task performance [Bor19]. Gómez Maureira and Verbeek [Góm16] argue that this might be because either the HMDs prevent an increased social presence, the HMDs increase social presence to be independent of physical proximity, or vocal communication has a higher impact on social presence than a visual representation of the other user.

2.3.1 Co-Located Collaboration

While collaborating remotely, users do not need to calibrate their VR or AR systems towards a shared coordinate system, because their relative positions in the real world do not matter. However, while engaging in a co-located collaboration users need to know where the other user is to avoid collisions [Hop18b, Oli18, Pod17, Pod18a, Rio18] or to be able to work together using tools [Sal09]. There exist several approaches that support the calibration of two co-located VR or AR devices. Billinghurst et al. [Bil00] and Kalkusch et al. [Kal02] use devices with cameras and fiducial markers placed around the room to calibrate the systems towards a shared coordinate system relative to the room. Different patterns can be used for marker detection [Sai07]. Modern AR frameworks allow to track arbitrary images as well as 3D models[21]. Bai et al. [Bai17] use this technique to integrate an HTC Vive controller interaction with a Microsoft HoloLens[22] HMD. Weissker et al. [Wei20b] calibrate two HTC Vive systems by using two trackers (one per user) with a known spatial relation. SynchronizAR [Huo18] uses an ultra-wideband (UWB) module, wifi, and a Google Tango[23] device. The system measures the position of each device and the distance between the

---

[21] https://developers.google.com/ar/develop/c/augmented-images/
https://library.vuforia.com/features/images/image-targets.html

[22] https://www.microsoft.com/hololens/

[23] https://developers.google.com/project-tango/

different devices. An algorithm minimizes the euclidean transformation and solves for the shared calibration of the devices. However, the method is imprecise, restricted to 2D planes and UWB modules are not available in any commercial off-the-shelf VR or AR product. Krohn et al. [Kro05] and Hazas et al. [Haz05] also calibrate different devices in a 2D plane using infrared light. Wahlstrom et al. [Wah16] calibrate a smartphone to a car using an unscented Kalman filter and the device's IMU. But, the precision allows only to determine the approximate location of the phone in the car and is not an accurate calibration. SLAM algorithms use a camera to reconstruct the environment of the device, e.g., [Tak17]. SLAM provides tracking to popular AR or VR systems such as HoloLens [Liu18] or Oculus Quest and Rift S[24]. They create a map of the environment and track the device with respect to this map using an RGB camera [Dao18]. Also, depth information can be included [Tak17]. By matching the generated environment maps of the different devices, a shared calibration can be found.

Avoiding collisions with virtual agents [Soh17] or remotely connected users [Rei20] can supply a more realistic and comfortable experience and maintain the feeling of co-presence [Rei20]. However, users can also collide with physical objects. Semi-immersive CAVE-based or AR systems allow the user to experience virtual content and see the real environment including her/his own body and the body of other co-located users, e.g., [Bec13]. This allows users to avoid collisions with the physical environment and with each other. However, full-immersive VR systems mask the real world completely. Therefore, mechanisms need to be provided that protect users from any physical harm.

Static room barriers [Cir09] or virtual companions [Cir12] can warn users about the limits of their physical room. Additional depth information can help to visualize real-world obstacles [Hua18a, Woz18, Woz20]. Some systems even try to actively avoid collisions by activating users' muscles using electrical muscle stimulation [Fal20].

---

[24] https://ai.facebook.com/blog/powered-by-ai-oculus-insight/

The risk of collisions with other co-located persons changes users' behavior. Users give anthropomorphic avatars more space than inanimate objects [San15]. One can observe that users walk more carefully around co-located users compared to remote users [Pod18a]. Furthermore, they are slowed down when avoiding collisions in VR due to more careful movement [Rio18]. The risk of collision is accurately anticipated but with delay [Oli18]. Besides the slower movement, users leave more clearance space to other co-located users compared to real-world or remote collaboration [Buc19, Pod18b]. Also, the effects of gender on the behavior of letting someone go ahead seems to be reduced when using virtual avatars [Buc19].

The appearance of the co-located users has an impact on the immersive experience and also on the capability to avoid collisions. Users react differently to varying avatar appearances (e.g., bounding shape, dummy, or realistic) depending on the scenario/task [Pod17]. They prefer more realistic avatars during explorative experiences and basic avatars for task-focused experiences. While bounding shapes are an effective way to decrease collisions, avatar visualizations or AR video see-through are faster and preferred [Sca17]. Avatars can also be visualized for other users that do not participate in the VE [Wil19].

Another approach to reduce collisions with other users is to not let users get too close to each other. Redirected walking and collision prediction can preemptively alter the user's paths of movement to avoid collisions. Existing algorithms prevent collisions for two [Azm17, Bac13], three [Don19a, Don19b, Don20], or more users [Bac19, Su07] even in nonconvex environments [Rob06].
Also, space separation can be achieved using interactions within the VE, e.g., returning a badminton ball to a specific area of the field so that the user moves over there [Mar18]. The disadvantage of the separation approach is that it does not aid collision avoidance when users are standing close to each other. Additionally, shared-space usage is preferred by users compared to tracking space separation [Lac17].

The above methods work well for multiple co-located users that experience a shared space CVE, i.e., virtual worlds that are aligned for all users. As users start exploring the VE, their physical space might be too small and a method for locomotion is required. Group teleportation methods [Wei19b, Wei20a] enable users to maintain the spatial relationship between them, even while exploring the VE. However, as soon as users start to explore the virtual worlds individually, the position of other co-located users and their respective virtual avatars do not match up anymore. Users risk colliding with the real human that they cannot see. In that case, group teleportation and avatar visualization methods can be combined to maintain collision avoidance capabilities. For this, additional ghost-like/shadow avatars [Lac17, Lan18] can be visualized that show the co-located user's real position next to their virtual avatar that is exploring the VE. Therefore, ghost-like avatars help to avoid collisions when the VEs of users are not aligned.

2.3.2 Remote Collaboration

Systems like the VideoWhiteboard [Tan91], ClearBoard [Ish92], and the ImmerseBoard [Hig15] allow two users to remotely collaborate on a digital whiteboard. The whiteboards display the drawings of the other user as well as her/his appearance using a camera. The ImmerseBoard from Higuchi et al. modifies the user's video to preserve eye gaze and gesture direction, intention, and level of agreement and thereby increases the sense of being together. The TeamWorkStation by Ishii et al. [Ish91] enables real-time collaboration by merging a desktop computer window and a camera-captured desk. The shared image and a video chat system allow two users to collaborate on a writing/drawing task. Matsukage et al. [Mat15] demonstrate a remote collaboration system that also connects two users via a video chat. Furthermore, the system transforms pointing gestures that one user makes on items, e.g., books, on her/his physical desktop to the respective items on the desk of the other user. This is achieved by extracting the hand region of the user from a camera and projecting it to the table surface of the other user with respect to the reference frames of the indicated items.

The advantage of using VR instead of a classical videoconference system is for example that remote face-to-face collaboration is enabled. Expressions can go beyond verbal explanations or 2D screen annotations. In VR, the users may use their whole body, just like in a real-life conversation. The remote collaboration allows faster reaction times and has no travel expenses since users can meet and work together in a virtual room without the need to be physically there [Gre17b]. Campbell et al. [Cam20] found significant improvements of VR over video communication regarding presence, closeness, and arousal.

On-site support is more efficient than a 2D video with audio because simple video fails to create an adequate shared visual workspace and further feedback needs to be provided [Fus00]. Vertegaal [Ver99] underlines the importance of communicating nonverbal cues, like a user's gaze, in a multiparty collaboration. A virtual laser pointer, a context that is larger than the local user's current view, and stabilized annotations can help the remote expert and increase task performance [Gau14]. Yet, a simple pointing visualization, i.e., controllable active laser, is not sufficient [Fus03b, Kur04]. Pointing using a virtual hand is a viable solution for spatial input [Gen13, Sod13]. Visualizing the other users' field of view and their grasping behaviors can increase awareness and understanding of their actions [Fra99]. However, hand pointing only serves as a highlight of a certain area and not as a referencing device [Kra06]. Pointing accuracy is reduced at larger distances and may need to be accompanied by speech. Voice, hand gestures, line annotations, and other features allow users to provide instruction and guidance during physical tasks [Tho19]. Integrating the hands of the remote expert into the workspace of the worker offers a greater sense of co-presence and allows flexible gestures [Hua18b]. The method of Vishnu [Le 15, Le 16a] allows two users to take the same perspective in an AR-VR supportive task by standing in the same position. Besides the user's arms, there exists one additional pair of arms that originates from the user's shoulders and represents the movement of the respective other user. Vishnu is faster compared to an annotated 2D camera stream, increases the feeling of co-presence, and eases the mapping between the guide's instructions and the worker's interactions. Overlapping avatars of two users can also be used to guide movements and posture [Hoa16].

A live 360° camera video gives the remote expert an independent view from the on-site worker [Lee17, Lee18, Teo18] and can even provide the feeling of standing side-by-side [Cai18]. A 360° view can be reconstructed from a camera with a smaller field of view [Kas14] and stabilized to maintain orientation [Kas16]. However, a 360° camera does not allow the remote user to walk around. Several systems [Hua13b, Hua13c, Piu17a, Piu17b, Tec12] utilize a 3D reconstruction of a scene to enable remote collaboration. The 3D model can be combined with a 2D video to enable efficient and user-friendly collaboration [Lee20]. Scene reconstruction [Kom17] and adaptations [Kiy99] also allow a seamless transition between different perspectives.

2.3.3 Influence of Perspective on Task Performance and Collaboration

It is advantageous to provide each user with a view that is not upside down nor rotated in any way. Ferraro and Kella [Fer92] investigate reaction times for a decision if a word is correctly spelled or if it is derived from a real word, e.g., plant and blant. They find the reaction times increase in proportion to the rotation angle. Times grow by a factor of 1.1 for 60° rotation, 1.4 – 1.7 for 120°, and 1.4 – 1.5 for 180°. Graf et al. [Gra84] let subjects read aloud an upside-down and upright oriented text, and ask them to cross out typographical errors and to answer comprehension questions. They find that subjects have more false-positive word markings of incorrectly spelled words if the text is rotated. The results show that upright reading is about two times faster than upside-down reading and that it is about two times easier to detect errors in a text if it is normally oriented. Yet, the comprehension question answering performance is better with rotated text. Graf et al. conclude as a reason for this that subjects read the text more slowly and therefore more carefully than in the normal oriented scenario, and that they engage in an analysis at word-level. Byrne [Byr02] evaluates four different text presentations, i.e., horizontal, ‘marquee’-style (letters cascaded downward), and 90° rotated left or right. He finds that a normal orientation is about 1.5 – 1.9 times faster than the other orientations. Annett [Ann91] confirms the reading performance results and measures upside-down reading to be about three times slower than normal reading in a pilot study.

VR makes it easy to alter the user's world and apply personal changes to it, e.g., to enable upright content display. However, content modifications can also be achieved by other technologies: Systems like Lumisight [Mat04], UlteriorScape [Kak08], TaPS Widgets [Möl11] or Bounsight [Osh10], all use Lumisty films[25] to allow for personalized views, e.g., upright-rotated text for each user, on a tabletop surface. PiVOT [Kar12] extends this concept from 2D to 3D auto-stereoscopic surfaces and Permulin [Lis14] provides personalized views on a shared 3D display. AFreeCA [Mar10] additionally proposes shared/public objects and personal objects that are only accessible by one user. All of these systems allow users to work on personalized-private and/or shared information. However, while content can be adapted individually, an adaptation of the user's body to fix any spatial inconsistencies between users is impossible in the real world.

Orientation and perspective also play an important role in multi-user systems. During *grounding* [Cla91], users initially establish a shared understanding to make sure that "what has been said has been understood". Grounding is performed following the principle of the least collaborative effort and changes with the medium. For example, a user's distance or movement towards virtual objects helps to define the referential context during conversations and thereby can be used by others to disambiguate references [Ott02]. Having a shared perspective also supports grounding and is "useful for understanding the situation and providing effective guidance and instruction" [Nag15] and allows users to "communicate with a common understanding, i.e., a sense that they are watching the same situation" [Kas14]. Having a shared workspace increases performance [Fus00, Ger04], especially when tasks are visually complex or difficult to describe [Ger04]. In spatial work environments, users try to build a mental representation of the world from the perspective of the other party [Pou16]. Guides try to diminish the mental workload on the user that is guided by using more neutral utterances or those that require the guide to apply mental rotations to take the perspective of the manipulator. This takes time (varying with the degree of rotation) and increases the mental workload

---

[25] http://www.glazingenhancement.com/products_lumisty_view_control_film.html

of the guide. In comparison, the guided person uses more ego-centered utterances. Pouliquen et al. [Pou16] state that a guide could be supported by tools that allow her/him to take the other user's perspective. Galati et al. [Gal13, Gal15] show that those effects are not bound to VR and arise in a real-life scenario as well. Moreover, a shared perspective is more effective and preferred to an opposing view, which can result in left-right ambiguities [Fei18, Tan10]. Wang et al. [Wan20] enable a user to share the perspective from another user without having to change her/his position. They use view splicing to combine the view of a worker and a guide.

But, a shared view via a 2D or 360° video can also constrain the perspective of the remote user to the recorded frame or the location of the on-site worker respectively. 3D reconstructed scenes solve this but can be challenging to dynamically update and lack in detail [Piu19]. Also, a first-person video may induce motion sickness because of a shaking and uncontrollable moving camera [Ben02, Nag15].

In the real world, comfortable between-user distances, i.e., personal space distances, are estimated at 45 – 122 cm in a casual-personal relationship [Lit65]. Also, people prefer to work corner-to-corner (sitting at a 90° angle) to communicate rather than face-to-face (sitting across from each other) or side-by-side (sitting next to each other) [Fes50, Som59]. Experiments show that users maintain personal space bubbles around virtual avatars similar in size and shape compared to bubbles maintained around actual humans [Bai03]. Additionally, peripersonal space boundaries are consistent with those in the real world [Buc20]. As previous work shows, real-world experiences and skills can be transferred to virtual worlds [Hop18a, Rin10] and that social norms of gender, interpersonal distance, and eye gaze from the physical world transfer even into non-immersive CVEs [Yee07b]. Face-to-face interaction provides powerful cues for intention by visual and auditory information [Shi09]. A side-by-side visualization rather than having the same viewport can increase performance and the feeling of co-presence [Cai19]. Having different perspectives enables spatial partitioning and identification/awareness of others' focus and activities [Fus00, Tan10]. Kunert et al. [Kun20] show that multiple displays, i.e., table and wall display, can support seamless collaboration while providing different perspectives into the VE. Following the other person's eye gaze improves performance [Gar01], allows to resolve ambiguity [Han07], and can also increase the likelihood to speak [Ver02]. Four of the 20 most frequent communication behaviors are listening, asking questions, discussing, and sharing information [Key13] and workgroups commonly switch between guided and collaborative sequences [Hal00].

2.3.4 Avatar Adaptation

Redirected walking [Raz01] adapts the VE imperceptibly for the user to allow seemingly unlimited walking inside a constrained physical space. This concept was quickly transferred to object manipulation. By offsetting the hand movement of the user, haptic feedback can be provided [Abt18, Car16, Han18, Koh09, Koh10, Rie18, Sai18] with dynamic targets [Che17] and arbitrary shapes [Yan18, Zha18]. The redirected touching performs equally to unmodified haptic feedback [Koh12, Koh13]. Most of these techniques require a rather static scene and a not navigating user and are thereby not suitable for room-scale VR experiences. Furthermore, redirection can be used to improve the sense of presence [Azm16], increase the user's reach [Feu17a, Pou96], or provide a more comfortable interaction [Feu18, Mon17]. Stoakley [Sto95] and Bowman [Bow01c] found that users do not notice if their virtual hands are positioned 10 cm or rather 25 cm higher than their real hands. They used this insight to reduce user's fatigue. Some work indicates that this effect can be maintained at even greater distances and other directions [Feu18, Kal14] especially if the displacement is away from the user [Feu17b, Kil12b]. However, other research shows that the effect cannot be maintained if the offset is greater than 30 cm [Llo07, Nie17c]. Moreover, redirection techniques can be applied to the whole body of the user to convey slow-motion movement effects [Rie17] or underwater drag forces [Kan19] without breaking presence.

Besides single-user modifications, redirection can also be applied to multi-user setups. Roth et al. [Rot15, Rot18] adapt the other user's avatar to imitate one's movements to increase interpersonal understanding and rapport. They found that the modification was mostly undetected and did not impact the perception of communication. Piumsomboon et al. [Piu18] increase users' social presence and task performance by modifying the position, scale, and body pose of a remote user's avatar to ensure that her/his gestures are always inside the local user's field of view. In later work [Piu19], they display a miniature avatar at the location of a 360° camera to allow collaboration between an on-site AR and a remote VR user. Sousa et al. [Sou19] mirror the worker's representation and the 3D workspace to provide two users facing each other with an identical point of view. They found that especially instructors benefit from the shared perspective regarding task performance and user preference as less coordinate system conversions are needed. Lee and Hua [Lee11] investigate the effects of adapting the view of the virtual world per user. In a pointing task, two users can rotate the content on a table consistently or independently from one another. World inconsistencies are fixed by transforming the laser pointer interaction towards a highlight of objects that are indicated. The user study results show that an independent world view can lead to faster task execution, easier pointing, and higher user preference. But, the proposed method conveys less co-presence and the inconsistent view makes it harder to recognize the positions of objects at which their collaborators point at.

Users focus more on the task space and hands of remote users rather than a video showing their faces [Tan10]. Also, workspace consistency might be more important than people's representation during remote collaboration [Sou19]. Yet, providing a full-body avatar is important and comes close to face-to-face interaction [Smi18, Yam18] and can be an effective tool for communication [Dod11, Sch10]. However, when modifying the other user's avatar one needs to consider that avatar realism is based on appearance and behavior [Gar03]. A realistic appearance can lead to heightened expectations for behavioral realism [Sla02]. Co-presence is increased when the character's behavior is realistic [Zib18] and decreased when appearance and behavioral realism mismatch [Bai05]. Chen et al. [Che14a] found that when users are presented with avatar ambiguities, i.e., offset between real person and virtual avatar, that this ambiguity leads to perceptual conflicts and a reduction in task performance during collaboration. Moore et al. [Moo07] argue that free gesticulation and publicly visible actions, e.g., menu access, are important to offer better coordination between users (see also [Hin98]). For example, Xia et al. [Xia18a] use ghost-like copies of avatars or objects to support multi-user movement or to resolve conflicts during simultaneous object interaction.

2.4 Conclusion

In conclusion, VEs can present users with immersive experiences to explore spatial content. Different hardware systems for visualization, tracking, and controls are needed to enable interaction and collaboration. The visualization of content or avatars has a large impact on the perceived experience and changes the perception of things such as depth or presence.

Diverse input devices can exploit proprioception or haptic feedback to increase the usability and performance of systems and users. Familiar devices and 2D interfaces allow users to transfer real-world skills and knowledge into the VE. However, current systems often require extended setup, e.g., wearable trackers or expensive hardware, and are not easily movable. Additionally, the currently used interaction with PC-based software within VR, i.e., projecting the window content on a plane as-is and using a ray-casting technique, does not exhaust all possibilities to transfer existing applications into VEs.

VR systems connect co-located and remote users to enable collaboration regardless of their physical locations. Efficient cooperation demands communication channels and expressive tools such as speech or spatial gestures as intentional and unintentional interpersonal communication is frequently used. But, the impact of the user's physical location, i.e., the differences between co-located and remote collaboration, is not yet fully explored. Furthermore convenient local-space calibration for all 6DOF devices is missing.

A view on upright-oriented content benefits task performance. In addition to that, taking the same perspective as another user facilitates a common understanding and is more effective but might violate personal space distances. At the same time, a face-to-face configuration allows to facilitate cooperation and grounding and is preferred by users. VR allows to modify the experiences of users and adapt their interaction and social behavior using redirection methods. Yet, redirection is only sparsely employed in the context of multi-user collaboration.

This work aims to integrate real-world commodity PC hardware and available PC software into VEs to boost input performance and the functional range of VR applications. Furthermore, through the integration of a full-body avatar and local space calibration for all 6DOF devices, co-located and even remote collaboration is enabled. Finally, this research combines remote collaboration with multi-user social redirection techniques to create an inherent mutual understanding by providing users with an individual perspective that fits their requirements to increase users' task performance and social presence.

3 Transferring Real-World Input and Output into Virtual Reality

Current VR systems allow users to explore immersive spatial environments using stereoscopic displays and spatial input controls. The new input and output devices offer exciting and natural interaction with virtual content. However, the spatial experience has drawbacks such as reduced task performance [Bow01d]. Also, it requires specially designed applications that have a limited functional range. To provide a familiar and efficient input and to increase the functional range of VR applications, real-world hardware (see section 3.1) and software (see section 3.2) are integrated into VEs.
The content of this chapter is based on the following publications [Hop18a, Hop19, Hop20b, Hop20c].

3.1 Virtual Proxy Interaction

HMD or CAVE-based VR systems often provide users with new input devices such as controllers/magic wands or use hand gestures. However, the different systems all have their distinct input devices. For newer users, it is quite difficult to just *pull the trigger* or *press the grip button* of a controller since the button layouts are unfamiliar. The new devices require learning and can be slow. Also, the mid-air interaction often is physically exhausting and lacks haptic feedback. More natural interfaces are needed. Gesture interaction via a camera that is mounted on the HMD lets the user grab virtual objects and manipulate the VE effortlessly. But, haptic sensations are missing.

To resolve these issues, a virtual proxy interaction with common devices can be used. A virtual proxy is a digital representation of a real-world object [Rus97]. In the work presented below, a virtual 3D model of a keyboard and a tablet are superimposed over a real-world device with a similar shape. Furthermore, the hands of the user are visualized in the VE. This yields a familiar interaction with devices that provide haptic feedback. Other approaches build upon using a 3D point cloud to visualize real-world objects or the integration of an AR video see-through display into the VR view. The advantage of integrating a virtual proxy instead of a 3D point cloud or an AR video see-through display is that the visualization of the haptic devices can be easily manipulated and further extended, e.g., depending on the context of the application.

3.1.1 Virtual Keyboard

The keyboard provides an easy and efficient way to input text into a system. To provide users with a similar input device in VR, *qVRty* was designed (see Fig. 3.1). The system consists of a real-world keyboard and a virtual 3D model that matches the real-world keyboard. The real-world keyboard is tracked using the Lighthouse system and its virtual counterpart is placed at the same location in the VR world. Furthermore, qVRty integrates a Leap Motion[1] hand tracking system to detect the hand movements of the user. The hands are then displayed in VR. The location of the Leap Motion device is also tracked using the Lighthouse system. Since the keyboard is wireless it can be carried around the room and set up everywhere. However, the hand tracking solution is still cable bound and does not allow for usage too far away from the desktop computer (range can be increased with an extension cord). The virtual proxy of the keyboard is not an exact copy of the real-world version. Adjustments to the virtual keyboard contain a larger font size and a red highlight on button press. Additionally, the tracked hands of the user are represented as a skinny skeleton that decreases occlusion and allows a better view of the keyboard.

---

[1] https://developer.leapmotion.com/

Figure 3.1: View of the real keyboard (left) and its virtual counterpart (right).

3.1.1.1 Evaluation

Setup The virtual keyboard proxy was implemented using Unreal Engine 4 and an HTC Vive HMD.

Participants A within-subject user study with 13 participants (9 male, 4 female) was performed to evaluate the performance of the approach. Participants were M=31.8 years old (SD=12.7). On a scale from 0 (low) to 5 (high) users' experience with VR was M=1.769 (SD=0.927). On the same scale, subjects assessed their typing speed with M=3.308 (SD=0.630).

Tasks and Procedures Each user was asked to type three different texts with the three conditions: typing with a keyboard in real-life as a baseline (qVRty baseline), qVRty with hands (qVRty), and qVRty without the displayed hands (qVRty w/o hands). The order of conditions was randomized to compensate for training and fatigue effects in the combined results. Before the first round, users were given three minutes to accustom themselves to the keyboard. The approximately 1400-characters-long texts[2] contained letters, numbers, german special characters, and punctuation marks and were typed sentence by sentence.

---

[2] Animal's attribute texts for *Mammute*, *Blauwal*, *Mantelmöwe*, and *Tiger* from https://de.wikipedia.org/

Measurements The method of qVRty aims to leverage the haptic feedback of the physical device to enable easy typing. Ideally, users can transfer their real-world typing skills into the VE without any losses. Because of this, the following hypotheses were formulated for the results of the evaluation:

H_A Typing speed and errors of qVRty are not different from real-world typing in the baseline condition.

H_B Workload and usability of qVRty are not different from qVRty baseline.

H_C qVRty achieves better performance and ratings than qVRty w/o hands.

The performance of user's typing was measured regarding their speed in characters per minute (cpm) and total error rate [Sou01, Sou03]. Furthermore, RTLX and UEQ ratings were captured. To assess issues with hand tracking, the tracking availability was also measured.

3.1.1.2 Results

Quantitative Results The evaluated qVRty techniques are fast in comparison to the related work (see Fig. 3.2). The participants achieved a typing speed of M=218.698 cpm (SD=39.573) with the baseline of qVRty, M=156.169 cpm (SD=71.590) with qVRty, and M=160.899 cpm (SD=65.732) with qVRty w/o hands. Mauchly's test did not indicate any violation of sphericity ($\chi^2(2)=4.312$, p=.116). A repeated-measures ANOVA for the speed of users yields a significant difference between the conditions (F(2,24)=28.196, p<.001, N=13) with a large effect size (partial η^2=.701). Pairwise Bonferroni-corrected tests show a significant difference between baseline to qVRty and baseline to qVRty w/o hands (both $p \leq .001$).

Although users only performed one session with each technique, the error rates are average for a keyboard-based approach (see Fig. 3.3). Users achieved total error rates of M=6.6% (SD=3.5) with the qVRty baseline, M=12.0% (SD=5.1) with qVRty, and M=13.5% (SD=6.5) with qVRty w/o hands. The corrected error rates are M=6.2% (SD=3.5) with the qVRty baseline, M=11.3% (SD=5.3) with qVRty, and M=12.7% (SD=6.5) with qVRty w/o

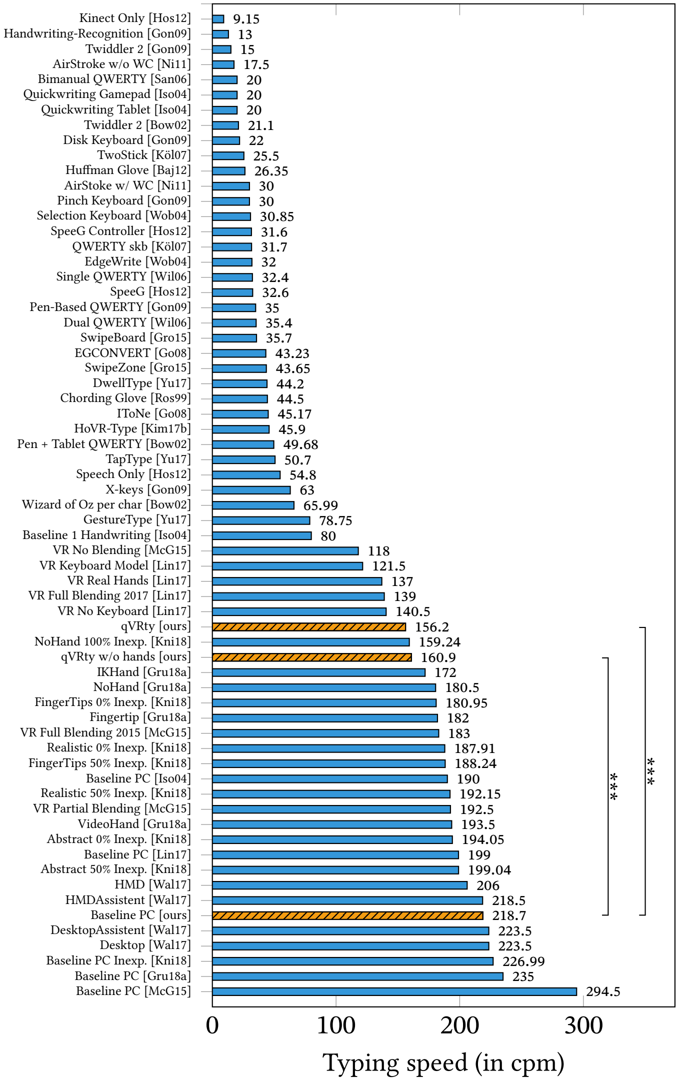


Figure 3.2: Input speed of different techniques in characters per minute (cpm). Values that were presented as words per minute (wpm) were converted to cpm by multiplication with five [Mac02].

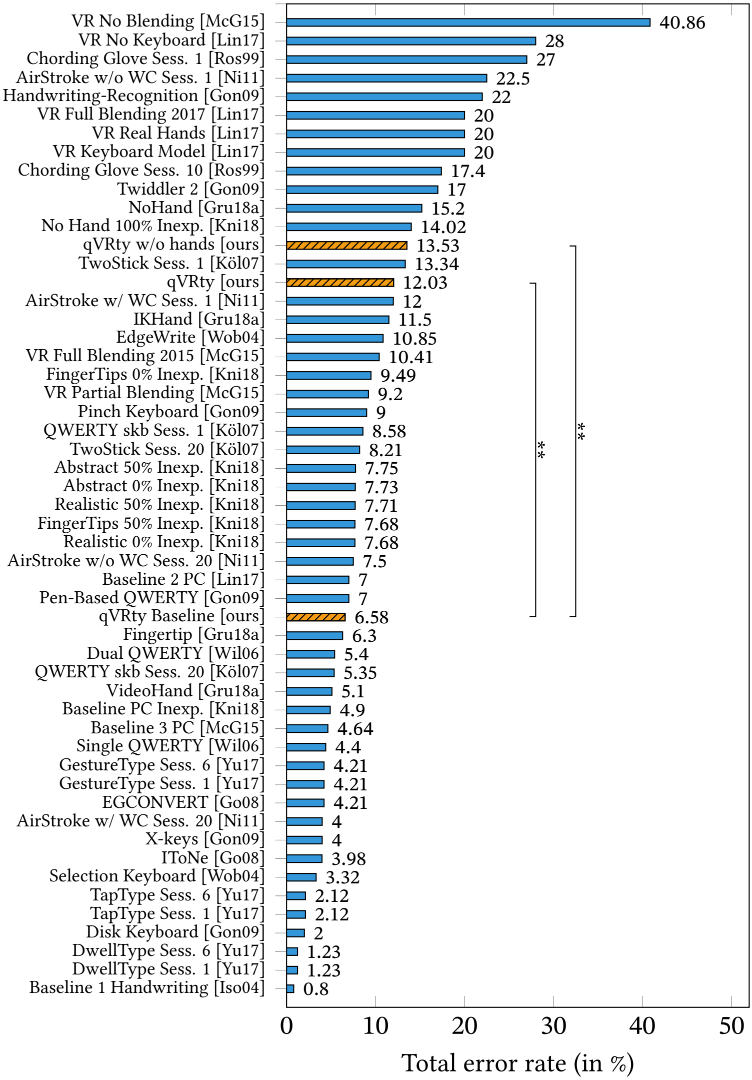


Figure 3.3: Total error rate [Sou01, Sou03] of different techniques in percent. If the total error rate is not available the corrected error rate is used since it is an estimation downward (value is always lower than total error rate).

hands. Mauchly's test indicates a violation of sphericity (χ^2(2)=9.515, p=.009, Greenhouse-Geisser ϵ=.633). A Greenhouse-Geisser-corrected repeated-measures ANOVA for the total error rate of users yields a significant difference between the conditions (F(1.267,15.200)=15.997, p=.001, N=13) with a large effect size (partial η^2=.571). Pairwise Bonferroni-corrected tests show a significant difference between baseline to qVRty and baseline to qVRty w/o hands (both p$\leq$.004).

Because of some tracking issues, the hands of the users were not displayed the whole time in the qVRty condition. On average the hands were displayed M=83.3% (SD=16.4) of the time. The minimum and maximum display ratio was 58.6% and 99.8%.

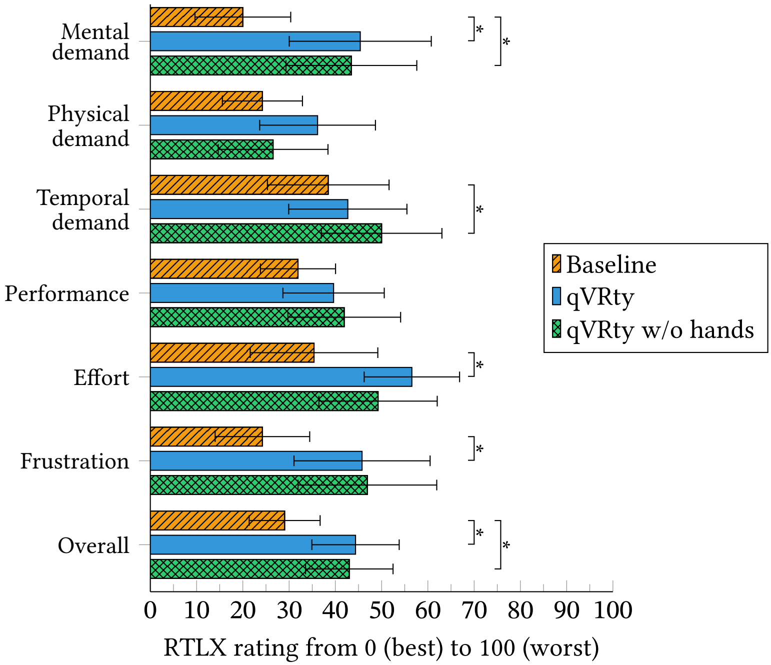


Figure 3.4: RTLX scores. Results as average value with 95% confidence intervals and asterisks indicating significant differences.

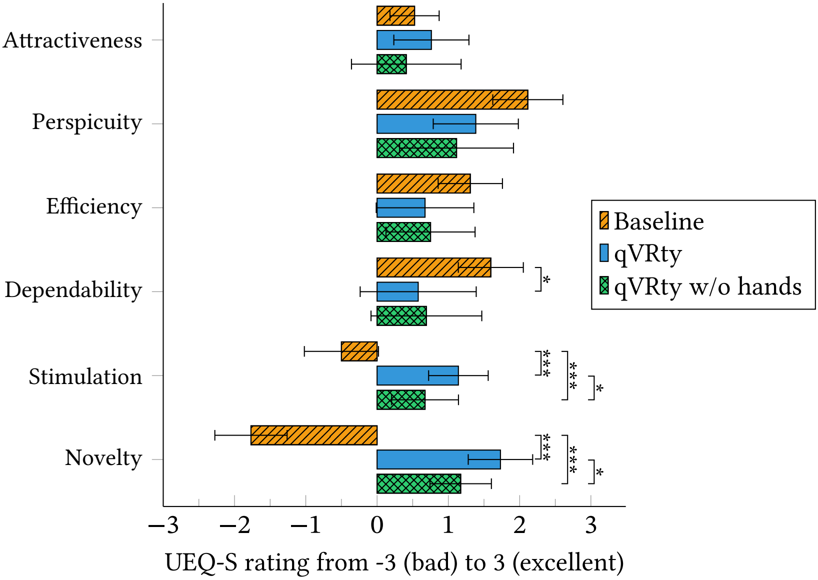


Figure 3.5: UEQ scores. Results as average value with 95% confidence intervals and asterisks indicating significant differences.

Qualitative Results According to RTLX and UEQ (see Fig. 3.4 and 3.5), typing with a virtual proxy is appealing and only slightly different from the baseline condition. On the RTLX subscale of mental demand, users rate the baseline condition with M=20.000 (SD=19.039) lower than qVRty M=45.385 (SD=28.245) and qVRty w/o hands M=43.462 (SD=26.012). A Friedman test shows significant differences for the mental demand ($\chi^2(2)$=13.220, p=.001, N=13) with a large effect size (Kendall's W=.508). Pairwise Bonferroni-corrected tests show a significant difference between baseline to qVRty (p=.018) and baseline to qVRty w/o hands (p=.013). The physical demand of typing is M=24.231 (SD=15.922) with baseline, M=36.154 (SD=23.018) with qVRty, and M=26.538 (SD=21.831) qVRty w/o hands according to users. A Friedman test shows no significant differences of the conditions ($\chi^2(2)$=3.368, p=.186, N=13). The temporal demand of the task was rated as M=38.462 (SD=24.185) for baseline, M=42.692 (SD=23.507) for qVRty, and M=50.000 (SD=23.979) for qVRty w/o hands. A Friedman test shows significant differences for the temporal demand ($\chi^2(2)$=9.784, p=.008, N=13) with a medium effect size (Kendall's W=.376). Pairwise Bonferroni-corrected tests show a significant difference between baseline to qVRty w/o hands (p=.032). Users rate their performance as above average for baseline M=31.923 (SD=14.936), qVRty M=39.615 (SD=20.152), and qVRty w/o hands M=41.923 (SD=22.411). A Friedman test shows no significant differences of the conditions ($\chi^2(2)$=1.955, p=.376, N=13). Users assessed their effort as M=35.385 (SD=25.369) for baseline, M=56.538 (SD=18.972) for qVRty, and M=49.231 (SD=23.527) for qVRty w/o hands. A Friedman test shows significant differences for the effort ($\chi^2(2)$=8.933, p=.011, N=13) with a medium effect size (Kendall's W=.344). Pairwise Bonferroni-corrected tests show a significant difference between baseline and qVRty (p=.024). The frustration of users is medium-low for baseline M=24.231 (SD=18.802), and medium for qVRty M=45.769 (SD=27.068) and qVRty w/o hands M=46.923 (SD=27.579). A Friedman test shows significant differences for the frustration ($\chi^2(2)$=7.702, p=.021, N=13) with a small effect size (W=.296). Pairwise Bonferroni-corrected tests show a significant difference between baseline and qVRty (p=.032). Finally, the overall workload of the three conditions can be derived. The baseline condition M=29.038 (SD=14.109) has a lower workload than qVRty M=44.359 (SD=17.368) and qVRty w/o hands M=43.013 (SD=17.396). A Friedman test shows significant differences for the conditions ($\chi^2(2)$=9.385, p=.009, N=13) with a medium effect size (Kendall's W=.361). Pairwise Bonferroni-corrected tests show a significant difference between baseline to qVRty (p=.018) and baseline to qVRty w/o hands (p=.032).

Users ranked the attractiveness of their experience using the UEQ as M=0.526 (SD=0.634) for baseline, M=0.762 (SD=0.969) for qVRty, and M=0.410 (SD=1.415) for qVRty w/o hands. Mauchly's test did not indicate any violation of sphericity ($\chi^2(2)$=2.435, p=.296). A repeated-measures ANOVA for the attractiveness shows no significant difference between the conditions (F(2,24)=0.741, p=.487, N=13). The perspicuity of conditions was ranked as M=2.115 (SD=0.905) for baseline, M=1.385 (SD=1.098) for qVRty, and M=1.115 (SD=1.471) for qVRty w/o hands. Mauchly's test did not indicate any violation of sphericity ($\chi^2(2)$=1.918, p=.383). A repeated-measures ANOVA for the perspicuity yields a significant difference between the conditions (F(2,24)=5.130, p=.014, N=13) with a large effect size (partial η^2=.299). However, pairwise Bonferroni-corrected tests show no significant differences. The efficiency for the baseline condition yielded M=1.308 (SD=0.830). qVRty and qVRty w/o hands achieved M=0.673 (SD=1.260) and M=0.750 (SD=1.150) respectively. Mauchly's test did not indicate any violation of sphericity (χ^2(2)=1.013, p=.603). A repeated-measures ANOVA for the efficiency shows no significant difference between the conditions (F(2,24)=2.610, p=.094, N=13). The baseline condition has high dependability with M=1.596 (SD=0.839). The average of the other two conditions is lower: qVRty M=0.577 (SD=1.498) and qVRty w/o hands M=0.692 (SD=1.429). Mauchly's test did not indicate any violation of sphericity (χ^2(2)=0.279, p=.870). A repeated-measures ANOVA for the dependability yields a significant difference between the conditions (F(2,24)=4.967, p=.016, N=13) with a large effect size (partial η^2=.293). Pairwise Bonferroni-corrected tests show a significant difference between baseline and qVRty (p=.026). The stimulation of the task is low with baseline M=-0.500 (SD=0.957). qVRty M=1.141 (SD=.769) and qVRty w/o hands M=.673 (SD=.862) achieve higher ratings. Mauchly's test did not indicate any violation of sphericity (χ^2(2)=4.014, p=.134). A repeated-measures ANOVA for the stimulation yields a significant difference between the conditions (F(2,24)=49.480, p<.001, N=13) with a large effect size (partial η^2=.805). Pairwise Bonferroni-corrected tests show significant differences between baseline to qVRty (p<.001), baseline to qVRty w/o hands (p<.001), and both qVRty methods (p=.016). Similarly, the novelty of the conditions is low with the baseline condition M=-1.769 (SD=0.932) and high with qVRty M=1.731 (SD=0.832) and with qVRty w/o hands M=1.173 (SD=0.793). Mauchly's test indicates a violation of sphericity (χ^2(2)=9.543, p=.008, Greenhouse-Geisser ϵ=.633). A Greenhouse-Geisser-corrected repeated-measures ANOVA for the novelty yields a significant difference between the conditions (F(1.266,15.190)=64.381, p<.001, N=13) with a large effect size (partial η^2=.843). Pairwise Bonferroni-corrected tests show significant differences between baseline to qVRty (p<.001), baseline to qVRty w/o hands (p<.001), and both qVRty methods (p=.040).

After the tasks, users were presented with a final questionnaire including four five-point Likert scale ratings from 1 (not at all) to 5 (very much). Overall, the virtual hands were perceived as helpful with M=3.077 (SD=1.382). While working in the qVRty w/o hands condition, users tend to miss the virtual hands with M=3.231 (SD=1.589). The virtual keyboard was rated as quite immersive M=3.538 (SD=1.561). Additionally, its haptic feedback was experienced as helpful M=4.385 (SD=1.325).

In an open feedback section, four users expressed that they could quickly accustom themselves to the virtual typing. However, seven users indicated that the hand tracking was imprecise, e.g., had positional offsets or the finger movement was not accurate. Furthermore, six users had problems with a blurry display, especially at the outer area of the FOV of the VR glasses.

3.1.1.3 Discussion

The hypothesis H_A cannot be supported by the results. Typing with the qVRty virtual proxy is quite fast and reaches 71% of the baseline speed using mobile, commodity hardware, and little training. However, there is still a significant difference to the real-world typing which is also present in the related work. The faster VR Partial Blending [McG15] reaches only 65% of the associated baseline. The related work of Grubert et al. [Gru18a] achieves higher speeds (75 − 85% of baseline) and comparable error rates using a more precise Optitrack tracking system. In a user study of Knierim et al. [Kni18], inexperienced users achieve 70 − 88% of baseline speed using a similar setup. However, both their methods are more intrusive and less mobile compared to qVRty as they require fiducial markers attached to the hands or a green screen.

The error rate of qVRty increases by 183 − 204% compared to the real-world baseline. Other methods with a lower error rate but higher speed than qVRty, use an AR video see-through display that shows the desk or a marker-based tracking system. As described before, these systems offer less mobility and are more intrusive. Related work also shows that using an error correction functionality allows to further decrease error rates and increase typing speed.

H_B also cannot be supported. The RTLX questionnaire shows that the workload of VR is significantly higher than with the non-VR technique. The reduced precision of qVRty induces a loss of perceived performance and increases frustration. However, both qVRty with and without hands offer a greater amount of stimulation and novelty.

The results of the user study show that the performance of qVRty with and without hands is not significantly different. This shows that the displayed hands were not as helpful as intended and H_C cannot be supported. Due to hand tracking issues, the efficiency might have suffered. Knierim et al. [Kni18] argue that based on their data, realistic-looking hands offer the best typing performance and highest presence, especially for inexperienced typists. In contrast, Grubert et al. [Gru18a] find that users have a higher preference and performance with rendered fingertips or an AR video see-through display than with displayed hands. Based on the results presented here and in related work, there are two aspects to consider: First, as argued by Grubert et al. [Gru18a], small discrepancies between the visualization of the virtual and real hands may induce an uncanny valley effect. The mismatch disturbs the interaction as users cannot rely as much on their virtual hands which leads to slower input and higher error rates. The differences result from hand pose estimations by the used Leap Motion sensor or the inverse kinematic used in [Gru18a] as well as tracking inaccuracies and delay. When hand poses are replicated correctly as in [Kni18], the effect is negligible and the virtual hands can even improve performance and usability. Second, the haptic feedback of the virtual proxy device seems to be able to compensate for a lacking hand visualization. Fast 10-finger touch typers do not need to look at the keyboard most of the time and therefore do not benefit from the displayed hands as much. Knierim et al. [Kni18] even show that the performance of experienced typists is not significantly different from real-world typing regardless of hand fidelity or transparency.

As a result, a fallback of including an AR video see-through display [Gru18a] of the keyboard and the user's hands is an acceptable alternative to the integration of a virtual proxy. However, the AR video see-through display also has significant shortcomings. The virtual proxy approach allows to modify or extend the displayed content to offer additional value. For example, a virtual proxy keyboard can include a word-by-word feedback display which increases performance through a higher focus on the typed content [Kim19a]. Additionally, the keyboard layout of the virtual model can be adjusted depending on the current context to boost user experience [Sch19, Wei09].

3.1.2 Virtual Tablet

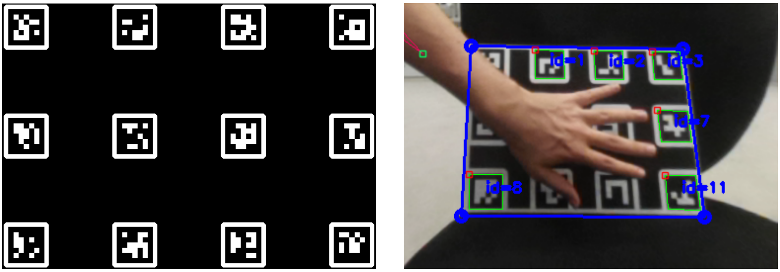
The keyboard is a highly efficient device for system interaction and is mostly used in desktop environments. In the real world, however, the most ubiquitous input method of current technology is touch. Smartphones, tablets, and even public information displays or vending machines use touch-based interaction as a means of easy-to-use and natural system input for users of all ages. To transfer the touch metaphor into VR, a virtual proxy of a touch tablet (VirtualTablet) was developed.

The VirtualTablet system was implemented on an HTC Vive Pro VR HMD. The Vive Pro features a 1440×1600 pixels display per eye with a 90 Hz refresh rate. The field of view of the Vive is 110°. Using the Lighthouse tracking system, the pose of the HMD is detected in a room area of 5×5 m^2. Moreover, the Vive has a stereo camera on the front of the HMD[3]. Each camera has a resolution of 640×480 pixels with a refresh rate of 60 Hz and a field of view of 96° horizontally and 80° vertically. The camera can be accessed with the SRWorks toolkit with a latency of about $100 - 200$ ms. SRWorks supplies depth information with an accuracy of ± 3 cm at 1 m and ± 10 cm at 2 m distance from the camera. The minimal working distance is 30 cm.

To detect the movable tablet, ArUco markers [Gar14] are used. They provide a position and rotation with respect to the camera, given their real-world size. The VirtualTablet consists of a simple rigid base, for example, a piece of acrylic glass or cardboard, pasted up with ArUco markers. Because the hand of the user will overlap with some markers while interacting, the markers are spread

---

[3] https://developer.vive.com/resources/knowledgebase/intro-vive-srworks-sdk/
https://www.stereolabs.com/blog/vive-pro-ar-zed-mini/



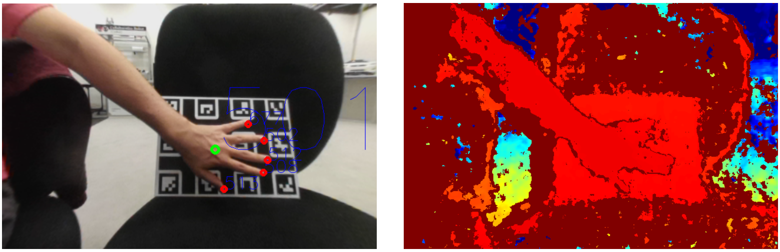
(a) ArUco board with 4×3 markers.

(b) Detected ArUco markers (green) and derived board pose (blue).

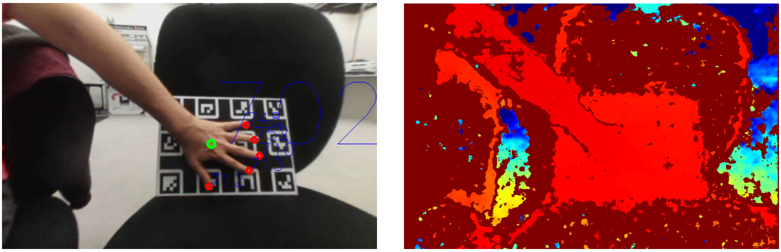
Figure 3.6: Setup and detection of the ArUco board.

out over the board (see Fig. 3.6). ArUco markers can be uniquely identified, therefore allowing the use of more than one VirtualTablet at the same time. The markers can also be hidden to increase aesthetics [Wil13].

To detect the fingertips, an RGB camera segmentation algorithm is used, since the hand of the user cannot be reconstructed from the given depth information (see Fig. 3.7). First, the hand is extracted from the RGB image by masking out the board. An interacting hand will be inside this area. Since the ArUco board only consists of black and white colors, these colors are subtracted from the remaining image which results in a binary representation of the segmented hand. Pixels with black or white color are removed in the HSV color space at a threshold with a low saturation (white) or a low brightness (black), i.e., $S \leq 80$ or $V \leq 40$ (scale from 0 to 255). As Fig. 3.8 shows, the binary image is then enhanced by a blur filter, the application of a threshold, and the extraction of larger surfaces, e.g., see [Bat12]. There are other ways to extract the hand by using skin color, so this approach can be easily exchanged if the background colors or lighting conditions change. Second, the fingertips are detected by calculating the extreme values of the contour of the hand [Che16b]. This is done by calculating the angle between three points on the contour that are 18 pixels apart from each other (see Fig. 3.9). To track the fingertips from one frame to another a distance-based tracking algorithm similar to [Xia18b] is used. Combined with the depth information from the stereo camera, the 2D tracked location of the user's fingers can be projected



(a) Floating hand: RGB camera. **(b)** Floating hand: Depth image.



(c) Resting hand: RGB camera. **(d)** Resting hand: Depth image.

Figure 3.7: Front camera images of a hand hovering above the board and touching the board. The resting hand cannot be distinguished from the tablet. The dark red color in the depth image marks invalid depth information.

into the 3D space relative to the HMD. To have a more robust depth value an average of 11×11 pixels is calculated around the 2D fingertip location. Since the pose of the surface and the fingertips are known, a touch can be detected. If the finger is closer than 0.5 cm to the surface, a click is triggered.

At the location of the real surface, a virtual tablet with the same size and shape is rendered (see Fig. 3.10). To give the user visual feedback of the location of her/his hand, a 3D point cloud is rendered on top of the tablet. White spheres at the fingertips highlight currently tracked fingers. The sphere turns green if a touch is detected.

The VirtualTablet can be used as a normal tablet which is familiar to many users. However, VR allows to enhance and extend the interaction with the tablet. The following paragraphs describe different techniques that are enabled by VirtualTablet.

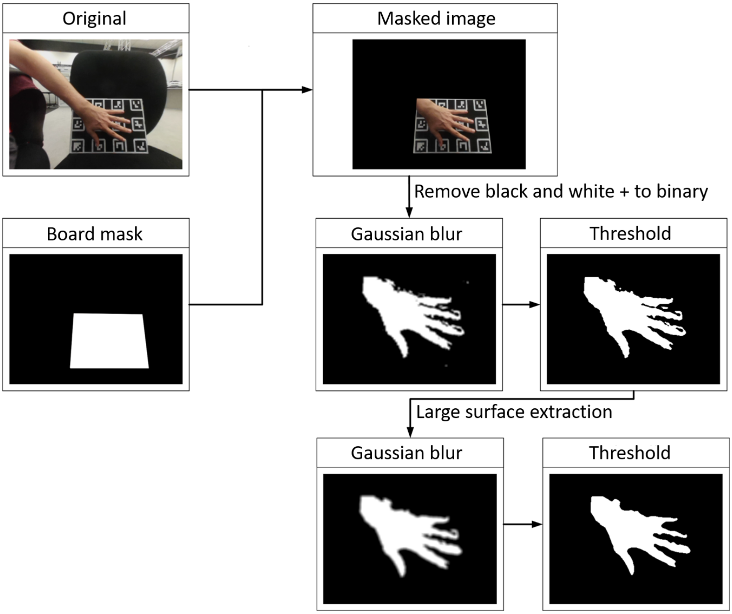


Figure 3.8: Hand segmentation process.

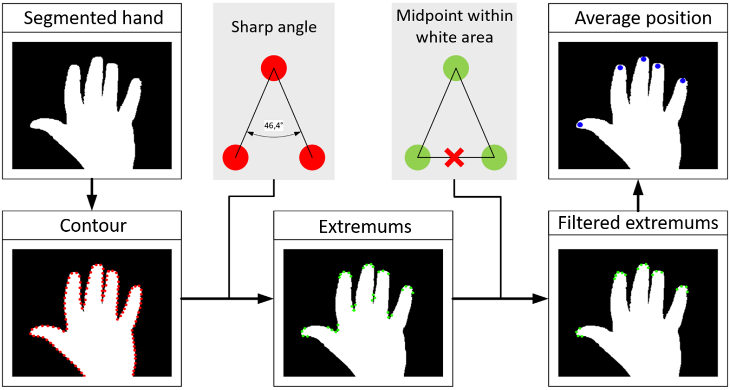
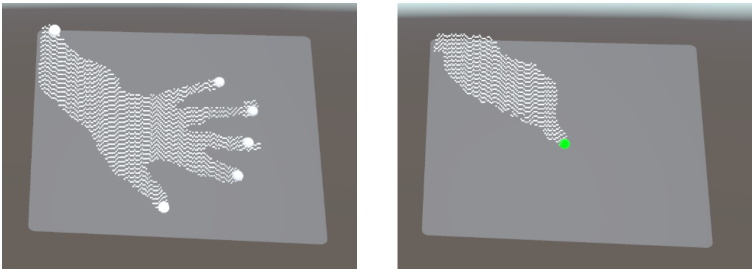


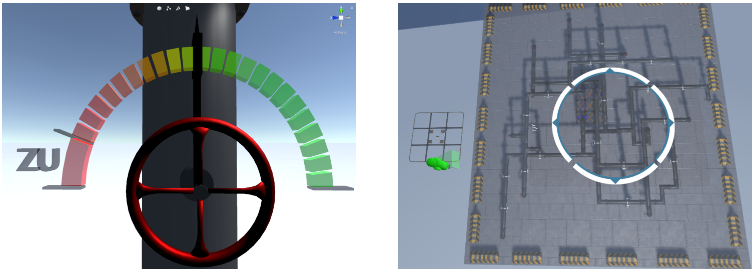
Figure 3.9: Fingertip detection process.



(a) Point cloud of the hand with detected fingertips (white spheres).

(b) Index finger touches the VirtualTablet as indicated by the green sphere.

Figure 3.10: Visualization of the VirtualTablet and the user's hand.



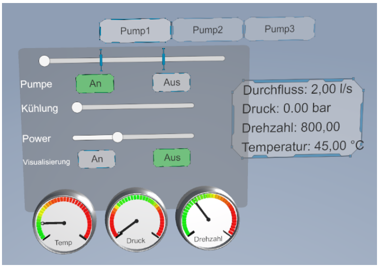
(a) A virtual valve wheel that can be turned by holding the VirtualTablet at the wheel's location and rotating it.

(b) The map of the VE can be moved by tilting the VirtualTablet. A green symbol represents the current location of the player. The user can teleport to a location by touching it.

Figure 3.11: Examples for a moving tablet interaction.

Even without touch detection, the VirtualTablet allows different forms of interaction. The position and rotation of the tablet can be used as an input to e.g., open a menu or change a value. If a user holds the tablet into a virtual object, an action can be triggered. By rotating the tablet at the location of a wheel, the wheel can be turned (see Fig. 3.11a). A navigation mechanism was implemented that allows the user to move a map around by tilting the tablet (see Fig. 3.11b). If the user touches a location she/he is teleported there (see [Bow98]).

Since the VirtualTablet has ArUco markers on both sides, it can recognize touch on any side of the tablet (if facing the user). This can be used to give the user a more natural way to navigate through menus. If the user flips the tablet around, a different UI is displayed. In the implemented application it is used to switch between an information interface and a map display for navigation.



Durchfluss: 2,00 l/s
Druck: 0.00 bar
Drehzahl: 800,00
Temperatur: 45,00 °C

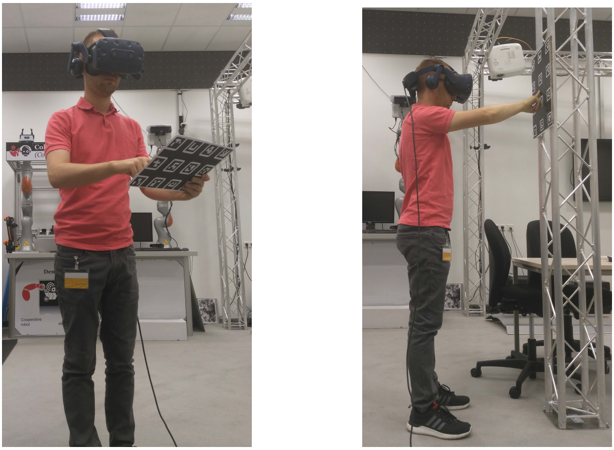
Figure 3.12: A menu that extends noninteractable information over the edges of the physical tablet.

VR allows to display information at any given location. In reality, displays are much more constrained and information cannot (yet) be displayed in mid-air. The VirtualTablet allows to extend the displayed content over the edges of the physical tablet (see Fig. 3.12 and [Miy18]). All interactable elements, like buttons and sliders, are displayed on the touchable surface. Other pieces of information are arranged around the tablet.

Physical tablets are expensive, rely on a power supply, and only come in distinct shapes and sizes. With this approach, multiple interactable surfaces can be created fast and cheaply. As seen in Fig. 3.13, a small mobile tablet and a larger mounted interaction surface were created. The different surfaces can be used in parallel.

3.1.2.1 Evaluation

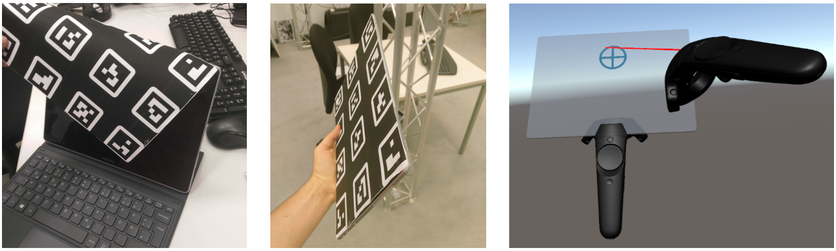
Setup To evaluate the proposed system and interaction methods, a within-subject user study was performed. An application was designed with Unity that allows a user to control different machines and valves in a factory using



(a) Mobile acrylic glass.

(b) Mounted cardboard.

Figure 3.13: Different interaction surfaces.



(a) Attaching the ArUco board to the capacitive touchscreen tablet.

(b) Holding the capacitive touchscreen tablet.

(c) Laser pointer and interaction surface attached to the HTC Vive controllers.

Figure 3.14: Other input techniques for comparison with the VirtualTablet.

the VirtualTablet. The task of the user was to react to a breach in a pipe system. While solving the issues, the user interacted repeatedly with the extended interaction methods described above.

To compare the detection performance of the VirtualTablet, three other input methods were implemented. A capacitive touchscreen tablet was equipped with an ArUco board (see Fig. 3.14a and 3.14b). The capacitive tablet uses the same hand visualization as the VirtualTablet but the touch detection from the display as a ground truth. The size of the tablet is about 1 cm smaller in width and 0.5 cm smaller in height than the acrylic glass VirtualTablet. The default interaction tools for the HTC Vive are the provided controllers. The pose of the controllers is tracked with sub-millimeter accuracy [Nie17a]. Applications often use a laser pointer to interact with a handheld menu (see Fig. 3.14c). This virtual pointer technique is effective and efficient [Pou98a]. Furthermore, an unmodified capacitive tablet in a non-VR scenario was used to evaluate the ground truth precision of touch interaction for the given tasks. The independent variables for the user study are the four described techniques VirtualTablet, CapacitiveTablet, and Controller (all in VR) as well as the Non-VRTablet. To compensate for the effects of learning and fatigue the conditions were counterbalanced.

Participants 28 people (24 male, 4 female) participated in the user study with an age distribution of 2 under 20 years, 13 from 21 to 30 years, 8 from 31 to 40 years, and 5 larger than 40 years. On a scale of 1 (none) to 7 (very much), users had M=3.214 (SD=1.934) experience with VR. All users had prior experience with a touch-based device and 15 users had used a tracked controller before.

Tasks and Procedures First, the users tested the extended interaction methods in the factory application. Second, the users were asked to perform several clicks and draw interactions with the tablet as in [Xia18b]. Users were presented with 6 targets which appeared four times each per technique. Each target had a size of 5×5 cm for all conditions. As a result, each user performed 6×4×4=96 clicks in total. The order of the click targets was random but precalculated for each technique so each user had the same sequence. For the shape tracing task, users were asked to trace a given figure (circle and line) as closely as possible and finish their drawing by clicking a button. Each shape was drawn twice.

Measurements The following hypotheses were formulated for the results of the evaluation:

H_A Input performance of the VR tablets is equal to or better than the controller interaction.

H_B Workload and usability of the tablet interaction are not different from Controller.

The precision of the target selection was measured by recording the position of the touch and calculating the distance to the center of the target. 49 outlier points (1.8%) with a distance of more than three standard deviations from the target point were removed as in [Har11, Xia18b]. Also, the time taken was measured. For the shape tracing tasks, the amount of filled target area was calculated. Furthermore, the percentage of pixels painted inside the image was measured. This value is an indication of how accurately a user could draw inside the target area. Moreover, the average distance of the touchpoint to the target shape without consideration of the width of the brush and the shape indicates how closely users could follow the target shape. Also, the duration of drawing was measured. To quantify the qualitative performance and experience of the different hardware setups, RTLX and UEQ-S questionnaires were used to assess workload and usability. These questionnaires were only collected for the VR techniques since the NonVRTablet is only used as ground truth for precision. At last, users were asked to rank the three VR techniques on a seven-point Likert scale regarding wearing comfort, quality of input, and an overall ranking.

3.1.2.2 Results

Quantitative Results The results of the selection tasks are listed in Fig. 3.15 and Table 3.1. The 95% confidence ellipses in Fig. 3.16 provide an overview of users' accuracy. The three VR input methods VirtualTablet, CapacitiveTablet, and Controller have a selection accuracy of about 5 − 7 mm. A robust TOST (rTOST) [Yue74] equivalence test was performed with an epsilon of 2.315. 2.315 mm is the uncleaned average distance of the ground truth tablet

interaction, meaning this value represents the accuracy of the users in a best-case scenario. The selection accuracy of VirtualTablet and CapacitiveTablet is equivalent with a small effect size ($p<.001$, Cohen's $d=.065$). Furthermore, the selection accuracy of VirtualTablet and Controller, and CapacitiveTablet and Controller is equivalent with a medium effect size with ($p=.003$, Cohen's $d=.412$) and ($p=.001$, Cohen's $d=.370$) respectively.
The accuracy of the six targets as seen in Table 3.1 reveals, that the VirtualTablet detection has a higher accuracy of more than 1 mm at the top row. The other techniques do not show such a large difference. Also, no learning effect can be detected in the collected data. Figure 3.17 shows the average distances to the center in a temporal order. The offset from the center stays approximately the same over the period of the selection task.

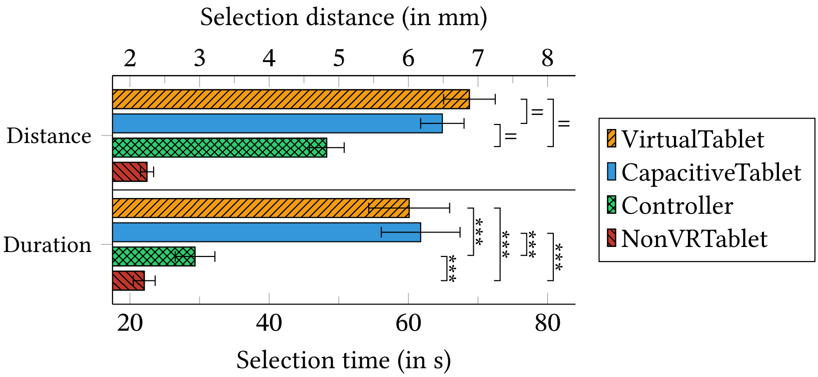


Figure 3.15: Selection distance and time for all 24 clicks. Results as average value with 95% confidence intervals, equal signs indicating significant equivalences, and asterisks highlighting significant differences.

The total amount of time needed to perform the selection task with each input technique is shown in Fig. 3.15 and Table 3.2. Users took about twice as much time with the VirtualTablet and CapacitiveTablet with regards to the Controller interaction and almost three times as much in comparison to the NonVRTablet. Mauchly's test indicates a violation of sphericity ($\chi^2(5)=35.726$, $p<.001$, Greenhouse-Geisser $\epsilon=.567$). A Greenhouse-Geisser-corrected

Table 3.1: Average selection distance (in mm). The 6 targets are coded as T/B for the top or bottom row and L/C/R for the left, center, or right column.

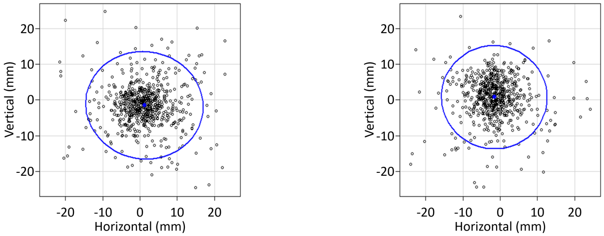
	Total	TL	TC	TR	BL	BC	BR
				VirtualTablet			
M	6.877	6.528	5.401	7.155	8.005	7.272	6.910
SD	4.869	4.701	3.603	4.827	5.287	5.287	4.985
				CapacitiveTablet			
M	6.488	6.553	6.491	7.218	6.752	5.877	6.023
SD	4.077	3.765	4.198	3.853	4.790	4.093	3.581
				Controller			
M	4.827	5.028	5.039	5.221	4.905	4.162	4.614
SD	3.291	3.445	2.929	3.619	3.345	2.783	3.501
				NonVRTablet			
M	2.246	2.527	2.118	2.301	2.562	2.159	1.820
SD	1.231	1.300	1.160	1.287	1.308	1.033	1.148

repeated-measures ANOVA for the duration yields a significant difference between the conditions (F(1.701,45.927)=84.910, p<.001, N=28) with a large effect size (partial η^2=.759). Pairwise Bonferroni-corrected tests show significant differences between the two VirtualTablet and CapacitiveTablet and the two Controller and NonVRTablet conditions (all p<.001). Additionally, the NonVRTablet is significantly faster than the Controller (p<.001).

Table 3.2: Average duration (in s) for 24 selections.

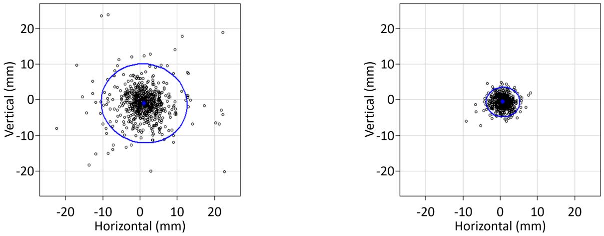
	VirtualTablet	CapacitiveTablet	Controller	NonVRTablet
M	60.120	61.770	29.350	22.060
SD	15.690	15.270	7.695	4.243

The results for the shape tracing task are shown in Table 3.3 and 3.4 for the circle and line target respectively. Drawing with the NonVRTablet achieves the highest tracing accuracy regarding the percentage of drawing inside the target area and target area filled. The CapacitiveTablet is on average slightly better



(a) VirtualTablet

(b) CapacitiveTablet



(c) Controller

(d) NonVRTablet

Figure 3.16: Selection offset from targets with 95% confidence ellipses.

Figure 3.17: Average distance of clicks chronologically ordered.

than the Controller and VirtualTablet condition. The standard deviation is smallest for the VirtualTablet and the NonVRTablet. The average distance towards the circle target area is very similar to this, with the NonVRTablet performing best with 1.9 mm accuracy and VirtualTablet performing worst with

a distance of 4.3 mm. For the line segment, the average distance of the CapacitiveTablet is very close to the NonVRTablet with 2.4 – 2.8 mm. The detected touches of the VirtualTablet have the highest average distance (4.3 mm) but are not so far away from the Controller tracing (3.4 mm). Figures 3.18 and 3.19 show the shape traces of all users on top of each other. The touch detection of the VirtualTablet leads to a lot of smaller errors outside of the target area. The ground truth touch of the CapacitiveTablet is a lot more stable. However, also the Controller drawings appear shaky. Finally, the NonVRTablet matches the target area the most.
Regarding shape tracing duration, the Controller is on average faster but more inaccurate than the CapacitiveTablet. The NonVRTablet is the quickest and the VirtualTablet takes the most time.

Table 3.3: Shape tracing accuracy for the circle.

	VirtualTablet	**CapacitiveTablet**	**Controller**	**NonVRTablet**
	Amount of drawing inside target area (in %)			
M	35.131	47.709	45.059	63.884
SD	8.884	11.847	11.239	8.949
	Amount of target area filled (in %)			
M	56.840	66.057	63.671	77.615
SD	14.452	16.490	17.694	13.091
	Average distance to target shape (in mm)			
M	4.274	2.960	2.892	1.873
SD	2.106	1.851	1.003	0.590
	Drawing duration (in s)			
M	15.688	8.128	6.455	5.277
SD	8.540	2.877	3.542	2.378

Qualitative Results Figure 3.20 shows the results of the RTLX questionnaire. Users rate the mental demand of the task as M=35.714 (SD=24.125) for the VirtualTablet, M=25.510 (SD=21.422) for the CapacitiveTablet, and M=19.558 (SD=17.063) for the Controller. A Friedman test shows significant

Table 3.4: Shape tracing accuracy for the line segment.

	VirtualTablet	CapacitiveTablet	Controller	NonVRTablet
	Amount of drawing inside target area (in %)			
M	39.058	56.813	49.715	82.004
SD	9.464	14.805	18.014	11.840
	Amount of target area filled (in %)			
M	61.949	67.469	64.202	78.607
SD	13.468	17.716	23.724	11.706
	Average distance to target shape (in mm)			
M	4.340	2.757	3.430	2.400
SD	1.383	1.387	1.640	0.848
	Drawing duration (in s)			
M	6.885	3.909	3.148	2.521
SD	4.075	1.424	2.179	1.172

(a) VirtualTablet **(b)** CapacitiveTablet **(c)** Controller **(d)** NonVRTablet

Figure 3.18: Stacked drawings of all circle traces.

(a) VirtualTablet **(b)** CapacitiveTablet **(c)** Controller **(d)** NonVRTablet

Figure 3.19: Stacked drawings of all line traces.

differences for the mental demand ($\chi^2(2)$=14.646, p=.001, N=28) with a small effect size (Kendall's W=.262). Pairwise Bonferroni-corrected tests show a significant difference between VirtualTablet and Controller (p=.001). The physical demand was rated as M=32.143 (SD=24.268) for the VirtualTablet, M=43.197 (SD=26.779) for the CapacitiveTablet, and M=25.850 (SD=22.087) for the Controller. A Friedman test shows significant differences ($\chi^2(2)$=8.970,

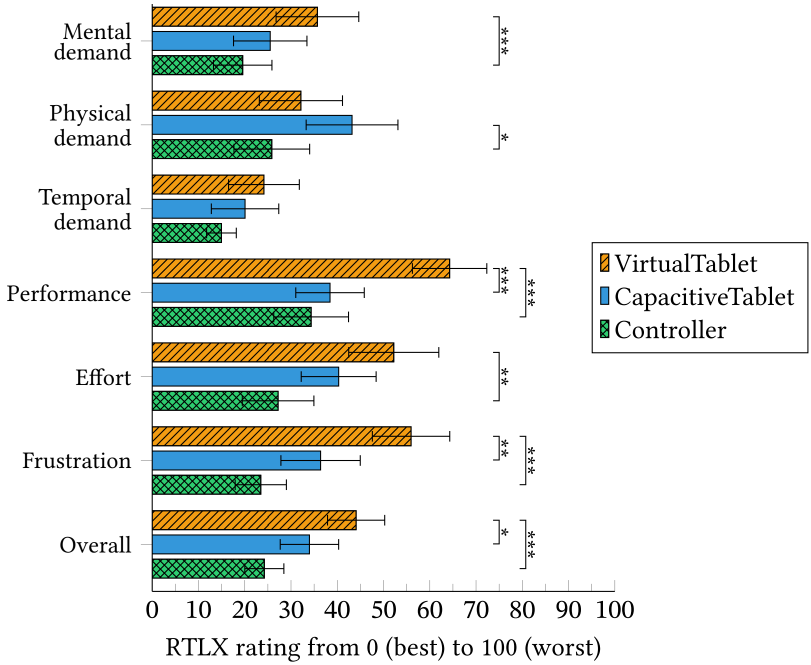


Figure 3.20: RTLX scores. Results as average value with 95% confidence intervals and asterisks indicating significant differences.

p=.011, N=28) with a small effect size (Kendall's W=.160). Pairwise Bonferroni-corrected tests show a significant difference between CapacitiveTablet and Controller (p=.015). As there was no time limit, users rank the temporal demand as low with M=24.150 (SD=20.652) for the VirtualTablet, M=20.068 (SD=19.670) for the CapacitiveTablet, and M=14.966 (SD=8.666) for the Controller. A Friedman test shows no significant differences for the conditions (χ^2(2)=1.194, p=.550, N=28). Users perceived performance is reduced while working with the VirtualTablet M=64.286 SD=(21.706) when compared to the CapacitiveTablet M=38.435 (SD=19.991) and Controller M=34.354 (SD=21.816). A Friedman test shows significant differences for the conditions (χ^2(2)=20.074, p<.001, N=28) with a medium effect size (Kendall's W=.358). Pairwise Bonferroni-corrected tests show that the Controller

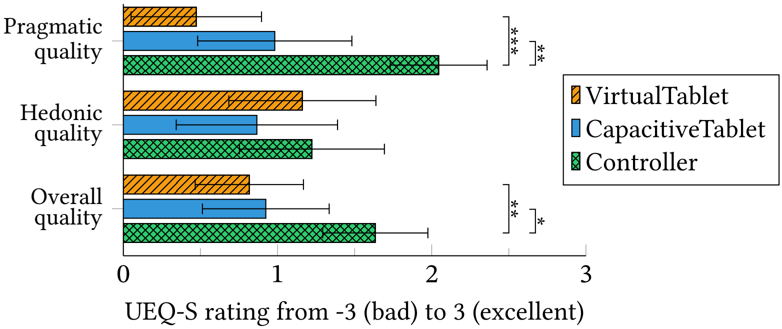


Figure 3.21: UEQ-S scores. Results as average value with 95% confidence intervals and asterisks indicating significant differences.

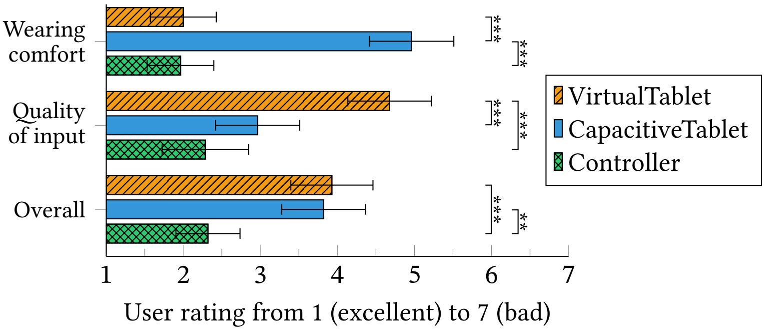


Figure 3.22: Rankings of the techniques. Results as average value with 95% confidence intervals and asterisks indicating significant differences.

and CapacitiveTablet have a significantly better performance than the VirtualTablet (both p≤.001). User's effort is rated as M=52.211 (SD=26.290) for the VirtualTablet, M=40.306 (SD=21.860) for the CapacitiveTablet, and M=27.211 (SD=20.892) for the Controller. A Friedman test shows significant differences regarding effort (χ^2(2)=10.000, p=.007, N=28) with a small effect size (Kendall's W=.179). Pairwise Bonferroni-corrected tests show a significant difference between VirtualTablet and Controller (p=.008). The frustration of users is higher with the VirtualTablet M=55.952 (SD=22.583) than with the CapacitiveTablet M=36.395 (SD=23.156) and the Controller

M=23.469 (SD=14.942). A Friedman test shows significant differences for the conditions (χ^2(2)=27.711, p<.001, N=28) with a medium effect size (Kendall's W=.495). Pairwise Bonferroni-corrected tests show that the Controller (p<.001) and CapacitiveTablet (p=.004) have a significantly lower frustration than the VirtualTablet. Finally, the overall workload is highest for the VirtualTablet M=44.076 (SD=16.683). The CapacitiveTablet and Controller are rated as M=33.985 (SD=17.047) and M=24.235 (SD=11.402) respectively. A Friedman test shows significant differences for the workload of the conditions (χ^2(2)=19.500, p<.001, N=28) with a medium effect size (Kendall's W=.348). Pairwise Bonferroni-corrected tests show that the Controller (p<.001) and CapacitiveTablet (p=.048) have a significantly lower workload than the VirtualTablet.

Figure 3.21 shows the results of the UEQ-S ratings. The pragmatic quality is bad for the VirtualTablet M=0.473 (SD=1.143), below average for the CapacitiveTablet M=0.982 (SD=1.350), and excellent for the Controller M=2.045 (SD=0.847) according to the UEQ Data Analysis Tool [Hin18a]. Mauchly's test did not indicate any violation of sphericity (χ^2(2)=0.719, p=.698). A repeated-measures ANOVA yields a significant difference between the conditions (F(2,54)=16.616, p<.001, N=28) with a large effect size (partial η^2=.381). Pairwise Bonferroni-corrected tests show a significant difference between VirtualTablet (p<.001) and CapacitiveTablet (p=.004) compared to Controller. The hedonic quality is rated above average for the VirtualTablet M=1.161 (SD=1.288) and Controllers M=1.223 (SD=1.268). The CapacitiveTablet's hedonic quality is below average M=0.866 (SD=1.412). Mauchly's test did not indicate any violation of sphericity (χ^2(2)=3.505, p=.173). A repeated-measures ANOVA shows no significant differences between the conditions (F(2,54)=0.970, p=.386, N=28). Finally, the VirtualTablet's and CapacitiveTablet's overall ratings are below average M=0.817 (SD=0.947) and M=0.924 (SD=1.110) respectively. The Controller is rated as excellent M=1.634 (SD=0.920). Mauchly's test did not indicate any violation of sphericity (χ^2(2)=0.489, p=.783). A repeated-measures ANOVA yields a significant difference between the conditions (F(2,54)=8.801, p<.001, N=28) with a large effect size (partial η^2=.246). Pairwise Bonferroni-corrected tests show a significant difference between VirtualTablet and Controller (p=.002), and CapacitiveTablet and Controller (p=.011).

Figure 3.22 shows additional ratings of the techniques. Users rank the wearing comfort of the VirtualTablet M=2.000 (SD=1.155) and the Controllers M=1.964 (SD=1.170) as good. The CapacitiveTablet is rated as below medium M=4.964 (SD=1.478). A Friedman test shows significant differences for the conditions ($\chi^2(2)$=34.750, p<.001, N=28) with a large effect size (Kendall's W=.621). Pairwise Bonferroni-corrected tests show a significantly higher wearing comfort of VirtualTablet and Controller when compared to CapacitiveTablet (both p<.001). The quality of input is ranked below medium for the VirtualTablet M=4.679 (SD=1.467) and good for the CapacitiveTablet M=2.964 (SD=1.478) and the Controllers M=2.286 (SD=1.512). A Friedman test shows significant differences for the conditions ($\chi^2(2)$=27.365, p<.001, N=28) with a medium effect size (Kendall's W=.489). Pairwise Bonferroni-corrected tests show a significantly higher input quality for CapacitiveTablet and Controller than for VirtualTablet (both p≤.001). Overall, the participants rank VirtualTablet M=3.929 (SD=1.438) and CapacitiveTablet M=3.821 (SD=1.467) as medium and the Controller as good M=2.321 (SD=1.124). A Friedman test shows significant differences for the conditions ($\chi^2(2)$=19.979, p<.001, N=28) with a medium effect size (Kendall's W=.357). Pairwise Bonferroni-corrected tests show that the Controller is rated significantly better than the VirtualTablet (p<.001) and the CapacitiveTablet (p=.005).

Before the performance-focused selection tasks, users were presented with some general questions regarding the interaction with the VirtualTablet during the explorative interaction. The custom questions were ranked on a seven-point Likert scale from 1 (best) to 7 (worst). The results show that the arrangement of the extended menus with information displayed outside the interactable area was clear M=2.857 (SD=1.597). Also, the practicality of having more than one input surface was rated highly M=2.429 (SD=1.116). Teleportation with the tilting map has a medium difficulty of M=4.107 (SD=1.718). During the user study users often needed help to initially understand what they needed to do. Yet, once learned, participants quickly understood the concept. Turning the valves with the orientation of the tablet and using the displayed buttons and sliders also has a medium difficulty of M=3.607 (SD=1.800) and M=3.643 (SD=1.315) respectively. The touch input and pose of the tablet and is recognized medium-well as users rate the expectancy as M=3.786 (SD=1.319) and M=2.571 (SD=1.400). Participants rate the minimum interaction distance of 30 cm as tolerable with M=2.929 (SD=1.534). Input delay was quite high (mostly due to the cameras), which is a little bit annoying for the users M=4.214 (SD=1.839). However, the haptic surface for hand gesture input is rated as helpful with M=1.500 (SD=1.086).

3.1.2.3 Discussion

The rTOST test shows that the VirtualTablet, CapacitiveTablet, and Controller perform equally accurately in a selection task. The precision of the VirtualTablet is about 6.9 mm, which is over half as small as the average size of a fingertip with 16 – 20 mm in diameter [Dan03]. The precision is good enough to interact with objects of suggested minimum target sizes, e.g., 9.6 mm [Par06] or 7 – 10 mm[4]. However, the shape tracing task shows that the touch detection of VirtualTablet is not robust enough. Due to a large amount of noise and invalid data in the calculated depth information, the touch detection is not able to always detect a continuous touch and sometimes detects false positives. The touch detection also fails if a user obscures the fingertip with her/his own hand. The speed of the VR touch systems is reduced by the large delay of the camera system. Besides increasing interaction times, this also affects the precision of the task execution, since the visualized hand and tablet locations do not match the current location of their real-world counterpart precisely. Because of this, H_A can only be supported partially. The precision loss and speed decrease influence the usability of the system negatively and lead to higher mental demand, effort, and frustration as well as less perceived performance and input quality. Therefore, H_B cannot be supported.

---

[4] https://support.google.com/accessibility/android/answer/7101858/

The results show that the VirtualTablet and CapacitiveTablet are twice as slow as the Controller input. Interaction with a controller differs from a touch input as users often only twist their wrist to point the laser towards another target. This is quicker than moving the whole hand on larger distances but it also induces more jitter as the stacked drawings show. The comparable MRTouch system [Xia18b] shows slightly better selection precision with M=5.4 mm (SD=3.2) when compared to the VirtualTablet. However, the VirtualTablet interaction uses movable touch surfaces, which induce additional precision errors. The tracing accuracy of VirtualTablet and MRTouch are comparable with an average distance of M=4.0 mm (SD=3.4) for MRTouch. Although the AR HMD of MRTouch uses a time of flight depth sensor and infrared cameras, the touch detection works with a threshold distance of 10 mm (VirtualTablet uses 5 mm) which could lead to the recognition of a touch before the user reached the surface. The very accurately tracked Controllers show a similar target selection result as VirtualTablet and MRTouch. This shows that touch input in VR and AR is very accurate, even with the present accuracy issues. The ground truth baseline NonVRTablet shows that there is room for improving the accuracy and speed of the presented technique to yield a touch interaction that is significantly better than a controller interaction.

VR does not allow the user to see her/his own hands. As with the virtual keyboard, accurate visualization of the user's hands seems to be crucial and has a larger impact on the precision and usability of the input technique. AR systems, such as MRTouch [Xia18b], already show the real-world hands and surfaces (compare keyboard input via AR video see-through display [Gru18a]). The used point cloud and fingertip spheres depend on the segmentation of the hand. The blur during head movements and the automatic brightness adjustments dampen the quality of the image. However, the point cloud representation by itself seems to be not so easy to understand spatially. The impact of the hand visualization in the shape tracing task is lower than in the selection task for the CapacitiveTablet because the user receives direct feedback from her/his touchpoint on the capacitive display and can compensate for any hand visualization delay or errors. Thus, the CapacitiveTablet performs better than the VirtualTablet.

The haptic surface for the interaction was rated as clearly helpful during the application task. This was also indicated through comments by the participants. Yet, as before, directly integrating touchscreen devices or showing the real-world hands and surfaces can have disadvantages. The extended functionality of a virtual proxy has benefits that a real-world touch tablet cannot offer, e.g., turning the tablet, making the hands or tablet transparent, or interacting above the surface [Men20]. Furthermore, The acrylic glass material has advantages over a regular touch device. Besides lower cost, no dependencies on electricity, and arbitrary shape and number of devices, the tablet is very lightweight (120 g) compared to the touchscreen tablet with a battery (754 g) and the Vive controllers (2×203 g). This leads to lower physical demand and higher wearing comfort, especially for longer sessions of usage.

3.2 Hybrid Interaction System

VR and AR experiences require specially designed applications that utilize stereoscopic rendering and often use gesture- or controller-based interaction. Yet, not all applications benefit from the advantages of a 3D representation and, on the contrary, are well suited for 2D visualization and interaction [Dar05], e.g., a web browser or spreadsheet application. However, a user might want to explore 3D data and then quickly browse the internet for information or check her/his calendar. It is not practical to implement a 3D version of each existing PC application. Also, users should not be forced to constantly switch between their 2D workspace and the 3D content by putting on and taking off the HMD. The 2D applications need to be accessible within the virtual worlds. There exist multiple ways to integrate 2D applications into a 3D world. Analog to the reality-virtuality-continuum by Milgram et al. [Mil95], interaction can either be fully real-world 2D, fully virtual 3D, or a hybrid of both. A 2D interaction can use an integration of an already existing application as a 2D canvas within the virtual world including a corresponding provision of input functions, e.g., left mouse button. Current PC-powered VR headsets allow the user to access the desktop via a 2D window and interact with it using a ray-casting interaction [Bow97b]. This approach is the simplest way of implementation, where all existing functions are immediately available. The synergy effects also save time by eliminating the administration of multiple implementations for developers. However, this approach has the disadvantage that a stereoscopic display and input are hardly or not at all used. As an alternative, the whole functionality can be implemented in 3D. With a 3D system, novel and possibly more natural interaction techniques, e.g., 3D grab manipulation, can be used for the operation of applications. Disadvantages of this variant are the increased effort for converting a 2D application to 3D and possibly a decreased interaction performance as spatial buttons and menus need more space and result in larger arm movements.

Both 2D and 3D applications offer various advantages and disadvantages. The proposed hybrid interaction system (HIS) combines 2D and 3D interaction and visualization. It targets to extend the functional range and increase the usability of VEs. An example implementation of the HIS is later presented in section 6.2. As a first step, existing 2D applications are integrated into the VE. To utilize the spatial visualization of VR, parts of the applications can then be transferred into 3D and synchronized with the remaining 2D content. To provide interaction with 2D and 3D content, a uniform mouse-based interaction system is proposed. Current approaches, such as ray-casting or handheld windows, integrate the virtual windows into the VE by capturing the 2D application as a texture and displaying it on a (curved) plane. The visual appearance of the 2D window is unchanged but only its position, rotation, or scale is transformed. To increase the usability of 2D applications that cannot be (partially) transferred to VR, different techniques for visualization and interaction are explored.

3.2.1 Integration of a 2D Application

To integrate any 2D application into the VE, a tool was developed which allows to capture the user's desktop and to send control inputs to the PC. The tool allows to display any 2D desktop window or a whole screen of the operating system inside a VR application. The desktop is captured using GDI (see section 2.2.3). GDI allows to address desktop windows via an identifier (handle) and to determine their device context. The device context defines the dimensions of the window. Based on this information, the window content can be stored in a separate texture (bitmap) by using the BitBlt (bit-block transfer) function. To enable interaction, the texture must be updated regularly. Typically, 30 frames per second (FPS) are sufficient. To allow the use of several 2D windows at the same time with a stable update rate, the image acquisition was parallelized. The image capture switches between two states: texture-update and texture-copy. During texture-update, a current screen/window is captured. During texture-copy, the new image is copied into the VR application and finally displayed on a 2D surface, e.g., a canvas, for the user. Since GDI does not capture the cursor, a cursor image is rendered at the respective location on top of the texture. The desktop capture was implemented with Windows 10 and Unity but can be extended to other operating systems and engines.

Additionally, a mechanism was implemented to forward control inputs, e.g., mouse cursor positions or keystrokes, to the desktop PC. Also, a ray-casting interaction was added. By calculating the intersection between the laser beam and the virtual window, the displayed position can be determined. However, this position corresponds to the world coordinate system of the VR application (XYZ, meters) and must first be converted into texture coordinates (UV, normalized 0-1) and then into desktop coordinates (XY, pixels). The UV coordinates can be extracted via the underlying game engine. The UV coordinates are then transformed into 2D application space coordinates by a simple bi-linear interpolation between the window's corners. With the help of the Win32 library[5], the position of the mouse cursor on the desktop PC can then be set correctly. The library also allows to transfer keystrokes (left mouse button, scroll wheel, ctrl key, etc.) to the VR application. As a result, any 2D application can be displayed and controlled within a VE.

---

[5] https://docs.microsoft.com/en-us/windows/win32/

3.2.2 Transfer 2D Content and Interaction into 3D

As mentioned, for some applications it makes sense to at least partially convert them into 3D, e.g., a CAD application with a 3D viewport and additional 2D menus. To convert a desktop application into a hybrid application, some modifications need to be made. First of all, the developer needs to identify which parts of the application are sensible to be visualized or interacted with within a VE. Generally, content that would naturally be presented as a 3D model, but is projected onto a 2D plane, is suitable for this. Examples for this are a 2D display of a car or a map which are just projections of a 3D CAD model or globe. Then, the respective content and interaction can be transferred into the VE with some implementation effort.

To enable communication between the original 2D application and the newly implemented 3D content, a message interface can be used. This allows to import and export information between applications and to synchronize the displayed content to reflect changes made in the realm of one to be also visible in the other. For this, each side implements a connection handler that can send and receive messages. The specific message queue implementation can use sockets, HTTP requests, event handlers, or other means to transmit data. The message itself should be of a general type to be extendible for different content and future changes. Therefore, JSON, XML, or similar formats are applicable. An example message can be found in Fig. 3.23. The message contains target and sender information which can be used to send the messages to the correct receiver. Additionally, as one target might receive different types of messages, the content type is specified. Lastly, the message may contain a content object, which specifies the arguments or any other information that is needed to synchronize the 2D application and 3D content.

3.2.3 Uniform Interaction with 2D and 3D Content

As users frequently switch between 2D and 3D content using the HIS, the interaction should follow a coherent concept. When users are not working within the 3D world, they will often be using a PC or a touch device such as

```
{
    "target": "<TargetName>",
    "sender": "<SenderName>",
    "content_type": "<ContentType>",
    "content":
    {
        "<AddtionalFields>": ...
        ...
    }
}
```

Figure 3.23: Example message with JSON syntax

a tablet. Both PCs and touch devices use a similar interaction metaphor, i.e., users can select (left click/touch) or scroll (scroll wheel/swipe) information. A right click/long press opens a context menu that allows to perform additional actions. To provide users with consistent and familiar interaction, a mouse-based interaction method is proposed to enable interaction within VEs for both 2D and 3D content. A controller mapping was implemented to support the following actions:

- Hover (no click): Ray-casting or touch collision
- Selection (left click): Index finger with trigger
- Context menu (right click): Thumb with button
- Scroll (mouse wheel): Thumb with touchpad or joystick
- Scroll press (mouse wheel press): Thumb with touchpad or joystick
- Modifier (Shift/Ctrl): Middle finger with grip button

To enable virtual objects to receive mouse-based input, an event-based interface was implemented. Virtual objects receive messages on action start, update, and end, e.g., on button press down, hold, and release. The mouse-based interaction is flexible and can be used to implement complex interactions within a VE. For example, if the user points on the ground and performs a left click, she/he may be teleported to the new location. If the user selects an object, a grab interaction may be performed. While holding the grab button, a scroll action may change the rotation of the object or its distance to the user. Also, a right click may open an additional menu that controls certain parameters of the object, e.g., its color.

3.2.4 Extension of the Visualization of 2D Applications

As some, but surely not all 2D applications can/should be (partially) converted to 3D using the proposed HIS, users will always need to interact with 2D content. Systems as described above or in the related work integrate 2D applications often as a (curved) plane into the VE which can be freely moved or is fixed to the body of the user or some device. Different ways are explored to find alternatives for a replacement of the PC mouse. However, to improve interaction performance, one also needs to consider the visualization of the content. Systems within the related work often include applications as-is or design applications specially for a certain use case. To facilitate a richer interaction with a 2D window in general, three different methods of optimizations were identified:

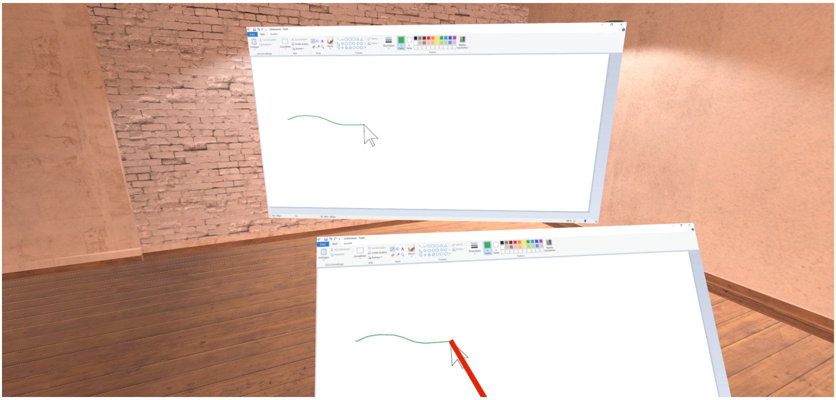


Figure 3.24: The user controls a copy of the captured window directly in front of her/him and sees a copy of the window for a better overview on the wall.

Users might quickly switch between a distant and a close-up, e.g., handheld or floating in-air, interaction with a movable virtual window. However, users need to move and scale the window to combine the benefits of having an overview over a large window and having a precise interaction with a handheld window. To easily enable both, a virtual window can switch between different discrete positions, i.e., snap to the wall or snap to the user's hands. Through snapping, a user's effort of manual positioning and scaling is reduced. However, since virtual windows can be duplicated at will, e.g., by cloning the virtual object, there is no need to switch between different window positions. The close-up and distant versions of a window can be displayed in parallel (see Fig. 3.24). This has the advantage that users do not need to rearrange one window but can work with different copies of it. In addition to that, several copies of one window can be used to implement multi-user interaction. A larger overview window can be used for group discussions and each user can work with her/his own close-up copy of the window. However, as users share the same application on one PC the interaction needs to be coordinated. The easiest solution to prevent users from interfering with each other is to integrate a locking functionality that allows only one user to interact with the window at a given time. Another approach is to have one single active cursor and several inactive cursors. The PC then only receives input from the active cursor. An inactive cursor can become active on a user's action, e.g., a button press.

Besides duplicating the whole window, users can also choose to duplicate only a specific part. For this, the user marks a specific area, e.g., using a rectangle selection. This allows the user to manipulate different sections of a window, like its toolbar or main content, individually. The various sub-windows can be scaled up and brought into a comfortable position close to the body to interact with them (see Fig. 3.25). This is especially useful to increase speed and precision for frequently used features or small menus or to extract and enlarge specific information. Also, this avoids the need to constantly change the size of a window, because some parts are too large while other parts are too small to access easily. This mechanism yields a tool that allows a user to build a workbench setup with the main content in the middle and all other important features in an easily accessible location.

Figure 3.25: The user extracted the save, redo and undo buttons, the brush panel, and the color panel from the original window to set up a customized workspace.

An extension of the user-controller duplication of window parts is an application-supported interaction. Analog to a mobile-mode for smartphones, applications can support a VR-mode. This VR-mode can include a flat visualization of the menu hierarchy and predefined sub-menus so that users can access any part of the application at will (e.g., see Fig. 6.5). Furthermore, complex menu workflows can be integrated using the proposed message interface between the 2D application and the VE. As opposed to displaying all sub-menus at once, different sub-menus can only become active if the user selects a certain function from a main menu or based on an event from the application. For example, a user may select a tool from the main menu. Next, a new sub-window containing additional settings for the tool may become active. The application may define default positions for the sub-menus that fit the content and interaction. In this example, the tool menu may be attached to the user's non-dominant hand. If the user activates a function that is not available, a pop-up message might be displayed in front of the user in conjunction with haptic feedback by the controllers. Moreover, users may detach, position, and transform sub-windows to build their individual workbench setup.

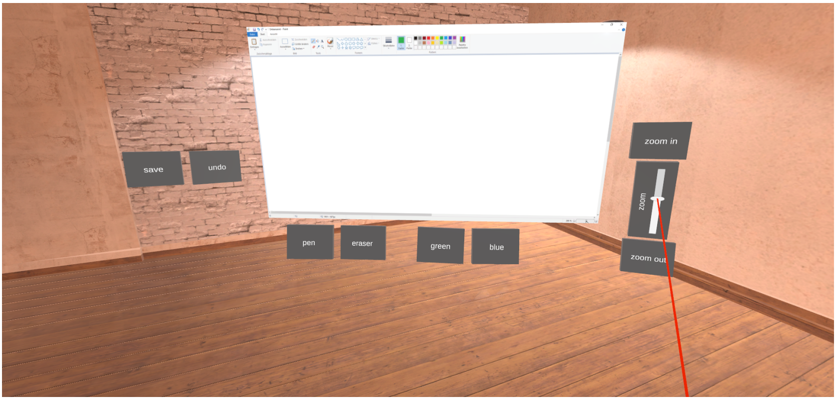


Figure 3.26: The user-defined macros for often-used features. The user is controlling a slider to change the zoom of the application.

Lastly, actions within the 2D application that require a mouse click/drag, keyboard button press, or any sequence of scriptable inputs can be automated. Users may create such scriptable actions, or macros, by programming a script, or by recording the actions directly in VR. The macros are represented inside the virtual world as a 3D widget (see Fig. 3.26). Once, the user presses the widget, the macro is executed. The macro tool can also be used to transfer common 2D UI widgets into 3D versions. A button or checkbox widget in VR could execute a simple macro that simulates a mouse click at a specific location. A 2D slider or scroll bar can be represented by a lever in 3D, where the start and end locations of the lever represent the start and end location of the slider in 2D. By moving the lever, a mouse click is simulated at the respective linearly interpolated location so that the 2D slider is set to the correct position. The macro tool allows easy and quick access to frequently used features and extends the interaction with the 2D window into the 3D world.

3.3 Conclusion

This chapter explored two approaches to increase the input performance and functional range of VR systems. The first approach uses virtual proxies to allow effective and easy interaction with familiar devices. The virtual proxies consist of a real-world haptic device and a virtual counterpart. The virtual proxy implementations include a keyboard and a tablet and support movable devices with a moving user, without wearable trackers, and with commodity hardware. Ideally, users can switch from the real world into the virtual world without experiencing any performance loss, taking their typing skills with them. But, the results of two user studies indicate that a precise integration of the user's hands is necessary to achieve real-world results. However, current precision seems to be sufficient to achieve acceptable results that are comparable to controller-based input, as the haptic feedback is already a great benefit. Furthermore, as the real-world devices are not integrated into the VE by, e.g., an AR video see-through display [Gru18a], it enables a variety of modifications and extensions of the input interaction.

The second approach integrates 2D applications into 3D VEs and targets to increase their ease-of-use and input effectiveness. It is unlikely that all existing 2D applications can be redesigned for usage in 3D. Nevertheless, access to standard programs is necessary to avoid frequent changes between the real and virtual world. Using screen capturing, arbitrary 2D applications can be displayed within a VE without reimplementation. If suitable, parts of an application can be transferred into 3D using the HIS and a message interface to synchronize the hybrid visualization. A mouse-based interaction system allows users to uniformly interact with both the 2D and 3D content. Furthermore, as not all applications might be fully transferable to VR, further tools are presented that extend the possibilities for interacting with virtual windows to provide quick and easy access to small menus or frequently used functions. These include the duplication and segmentation of virtual windows, the introduction of macros for quick access to certain commands such as copy and paste, and the proposal of a VR-mode that allows to integrate complex GUI interactions in VR with minor support from a 2D application.

When implementing interaction within VEs, one should consider the following insights and design recommendations:

VEs can build upon real-world skills: Virtual proxy devices in VR can provide input that is based on real-world skills. Two user studies show that usage of haptic virtual proxies achieves good performance and usability. Yet, even slight discrepancies due to tracking inaccuracies and camera delay reduce users' performance. Systems therefore should target to increase the precision of hand and device tracking to close this gap. Then, input in VR has the potential to become as fast as with real-world devices and faster than current controller-based input.

Partial integration of 2D applications: Interaction with 2D applications will make VR and AR technology more economically viable and increase user productivity by giving them quicker access to valuable information. Enabling the use of 2D applications increases the functional range of VEs strongly. However, when interacting with 2D, one should consider not only optimizing the interaction with content but the visualization of the content itself. If possible and applicable, an application may be partially transferred into 3D or made more accessible using a VR-mode to utilize the spatial representation and relationship of information and natural interaction with content.

Further advantages through enhancements of virtual content: One should consider to mimic real-world devices, e.g., using a virtual proxy, instead of integrating the devices themselves into the VE, e.g., through an AR video see-through display. The virtual representation of objects allows modifying their appearance based on the current context of the application, e.g., transparent hands or device, different button layouts. This increases their functionality beyond their real-world features. Furthermore, as the user does not need to interact with a functioning device, dummy objects, e.g., out of acrylic glass, can be used. These dummy objects are cheap, lightweight, and do not require a battery or data connection to the VR system.

4 Co-located and Remote Collaboration in Virtual Environments

Multi-user VEs allow several users to meet in a virtual room and collaboratively explore immersive visualizations. For this, users can work either co-located, at the same physical location, or remote, at separate physical locations.

On one hand, co-located collaboration has the advantages that users can talk directly and with no latency and supply haptic feedback to each other, e.g., exchange tools. However, when using HMDs, the different coordinate systems of users need to be calibrated to align the virtual worlds of all users. Therefore, an easy and precise calibration algorithm for 6DOF devices that is agnostic to operating systems and tracking techniques is presented below. While working in the same room, users might collide with each other, as their real-world awareness is strongly reduced. The requirements for a user's avatar visualization to prevent collisions are analyzed.

On the other hand, remote collaboration allows users to work together without the need for travel which saves time and resources. To determine the influences of the physical location of the user, a co-located and remote task execution is compared.

The content of this chapter is based on the following publications [Hop18b, Hop18c, Hop20a].

4.1 Co-Located Calibration for Arbitrary Six Degrees of Freedom Devices

Tracking algorithms are utilized to display spatial VR content that reacts to the movements of the user. Typically 6DOF, i.e., position and rotation, are captured to deliver an immersive experience. Besides tracking the head movement of the user, other body parts or input devices are tracked to enable interaction or realistic avatar poses. Modern VR HMDs use inside-out or outside-in tracking systems based on a variety of sensors, e.g., camera-/timing-/distance-based. Additionally, the different devices run on varying operating systems. The issue is, however, that tracking data is often not compatible between different devices or not even accessible by developers. No application exists that allows calibration of all available 6DOF devices. Therefore, a new method is proposed that registers the trajectories of two interlinked devices with 6DOF tracking. Through repetition with other devices, any number of systems can be calibrated towards a shared coordinate system. The algorithm has low requirements, as it only demands the position and angular velocity of each device as input which is the bare minimum sensory output that any 6DOF system offers. The position can be obtained through the 6DOF tracking and the angular velocity can be measured by an IMU or calculated by differentiating with respect to the changes in rotation.

In the beginning, two devices are interlinked (or held closely together) to form a rigid body and then moved around the room. During the movement, users should try to minimize any relative movement in between devices so that both devices form a rigid body system. Then, the rigid body system offers the following guarantees: (g_1) the distance of two arbitrary points is constant, and (g_2) the angular velocity ω at a certain point in time is identical for all points.

The algorithm uses these guarantees of the interlinked-bodies to calibrate two devices towards a shared coordinate system. Specifically, the algorithm calculates the rotation R and translation t between the coordinate systems of two devices (see Fig. 4.1). It consists of a time synchronization and mapping step and then finds the rotation offset and the translation offset of the two coordinate systems.

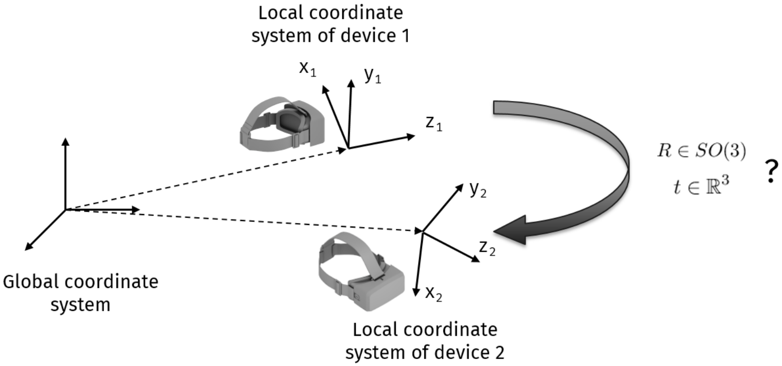


Figure 4.1: Two devices with 6DOF tracking have their own local coordinate system but lack a shared global coordinate system.

First, data recordings on both devices need to be triggered, e.g., by a message from a central server or one of the devices. The devices need to record data in the form of $(a_i) = ((p_{a_1},\omega_{a_1}),(p_{a_2},\omega_{a_2}),\dots,(p_{a_N},\omega_{a_N}))$ for *device A* and $(b_i) = ((p_{b_1},\omega_{b_1}),(p_{b_2},\omega_{b_2}),\dots,(p_{b_M},\omega_{b_M}))$ for *device B*, with p_{i_j} being the jth position of device i and ω_{i_j} the jth angular velocity of device i.

Resampling of the trajectory series allows the algorithm to calibrate two systems, even if they have different tracking rates. The position is resampled using linear interpolation $\hat{p}_j$ between two consecutive points p_j and p_{j+1} and is calculated as follows:

$$\hat{p}_j = p_j + \frac{p_{j+1} - p_j}{t_{j+1} - t_j} \cdot (\delta t \cdot i - t_j) \tag{4.1}$$

The angular velocity should be interpolated using spherical linear interpolation because its axis is not constant. However, as only the norm of the angular velocity is used in the following steps, it is interpolated only linearly:

$$||\hat{\omega}_j|| = ||\omega_j|| + \frac{||\omega_{j+1}|| - ||\omega_j||}{t_{j+1} - t_j} \cdot (\delta t \cdot i - t_j) \tag{4.2}$$

The tracking data is resampled with a frequency of f_s=90 Hz. This yields the series $(a'_i) = ((p'_{a_1},||\omega'_{a_1}||), \dots, (p'_{a_{N'}},||\omega'_{a_{N'}}||))$ as well as $(b'_i) = ((p'_{b_1},||\omega'_{b_1}||), \dots, (p'_{b_{M'}},||\omega'_{b_{M'}}||))$.

Yet, even though the resulting series have the same sampling rate, their starting time is not guaranteed to be time-synchronous due to network delay. One way to solve this is to synchronize the clocks of the two devices using a network time protocol [Cor05, Mil91]. However, this is not necessary. The presented approach uses the cross-correlation function [Rhu14] to synchronize the two trajectories by finding the maximum of the cross-correlation function of the angular velocity magnitude time series. This approach uses (g_2) and is based on the equation:

$$xCorr_{\omega'_a \omega'_b}(r) = \sum_{i=1}^{M'-r} ||\omega'_{a_i}|| \cdot ||\omega'_{b_{i+r}}|| \qquad \text{, w.l.o.g. let M' < N'} \tag{4.3}$$

Explicitly, the discrete series of $||\omega'_b||$ of the norms of angular velocities of *device B* is shifted over the series of $||\omega'_a||$ with an offset of r. The resulting sum reflects the similarity of the series at time offset r. To ensure that (b'_i) lies inside of (a'_i), a time interval of t_r=1 *s* is removed from the beginning and end of the data recording (b'_i) by setting the respective values to 0 which yields (b''_i). Because (b''_i) lies inside (a'_i) the factor $1/(M'-r)$ of the cross-correlation is not needed. By maximizing the rank correlation coefficient, the time offset Δt of (a'_i) and (b''_i) can be found using:

$$\Delta t = \frac{1}{f_s} \cdot \underset{1 \le r \le M'}{arg\ max}\ xCorr_{\omega'_a \omega''_b}(r) - t_r \tag{4.4}$$

The time offset Δt is then applied to the original series (b_i).

After the time synchronization, the series (a_i) and (b_i) start at the same time. However, their sampling rates might be different. Additionally, because data recording depends on application rendering times, the sampling intervals within both series might not be constant. To resolve any non-linear timing issues, a dynamic time warping (DTW) [Gol18] algorithm is used to receive

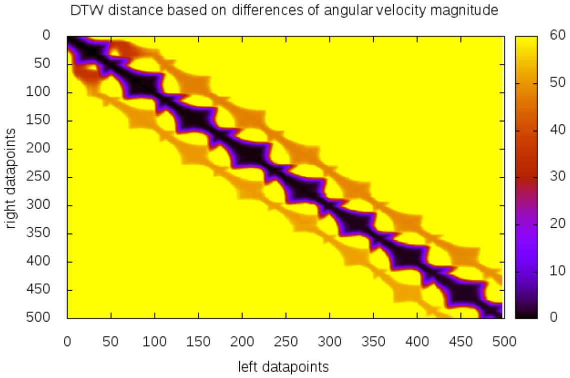


Figure 4.2: Dynamic time warping between angular velocity magnitudes of two headsets held together from a recording from the evaluation data set of fast rotation motions. In addition to the prominent main diagonal, two alternative diagonals are evident. These result from the periodic motion during recording.

a pairwise mapping between the series (a_i) and (b_i) (see Fig. 4.2). The calculations use the cost function $cost(x,y) = (||\omega_{a_x}|| - ||\omega_{b_y}||)^2$. DTW requires both series to start at the same time and end at the same time. Therefore the DTW algorithm's finish condition was adjusted to soften the requirement of the same ending time. Only pairs of diagonal steps within the warp path are kept in the series (a_i) and (b_i). Correspondingly, all pairs resulting from horizontal and vertical steps are discarded. The goal of this is to avoid that only one series progresses in time while data points of the other series are reused.

Based on the previous steps, the algorithm for the calibration of the devices can be defined. Data points in the series (a_i) are excluded if the distance between successive points p_{a_i} and $p_{a_{i+1}}$ is less than 1 cm. The respective mapped points in (b_i) are excluded as well.

Given the cleaned series (a_i) and (b_i), both of length M, with the same start time, the optimal rotation $\hat{R}$ between the coordinate systems of the two devices can be found by minimizing the angle between the axis for the pairs of angular velocities using a rotation offset R:

$$\hat{R} = \underset{R \in SO(3)}{arg\ min} \sum_{i=1}^{M} \measuredangle(\omega_{a_i}, R \cdot \omega_{b_i}) \tag{4.5}$$

Using the solved optimal rotation $\hat{R}$ and the guarantee (g_1), the optimal translation $\hat{t}$ can be determined. Because the distance between the devices is fixed, $\hat{t}$ can be found by minimizing the variance of the differences for the distances between the tracked positions points:

$$\hat{t} = \underset{t \in \mathbb{R}^3}{arg\ min} \frac{1}{M} \sum_{i=1}^{M} (d_i(t) - \tilde{d}_t)^2 \tag{4.6}$$

with $d_i(t) = p_{a_i} - (\hat{R} \cdot p_{b_i} + t)$ defined as the distance between the tracking positions of *device A* and the rotated positions of *device B* including a translation offset t, and $\tilde{d}_t$ the median of all distances for a translation offset t.

The concrete implementation of the presented algorithm uses COBYLA (constrained optimization by linear approximation) [Pow98], to optimize the functions $\hat{R}$ and $\hat{t}$. Also, unit quaternions are used instead of the rotation matrix to calculate $\hat{R}$. This decreases the number of parameters to optimize from nine to four. The unit quaternion $q = (x_q, y_q, z_q, w_q)$ is optimized under the constraint of $|1 - x_q^2 + y_q^2 + z_q^2 + w_q^2| < 0.01$. Initially, q is set to the identity quaternion, i.e., $x_q = y_q = z_q = 0, w_q = 1$.
Equation 4.6 is adjusted to avoid that COBYLA optimizes towards an unwanted local minimum:

$$\hat{t'} = \underset{t' \in \mathbb{R}^3}{arg\ min} \frac{1}{M} \sum_{i=1}^{M} (d_i(t') - \tilde{d_{t'}})^2 \cdot \frac{1}{M} \sum_{i=1}^{M} d_i(t') \tag{4.7}$$

The additional factor of the average point distance favors solutions where the distance between the two HMDs is smaller. The translation offset t' is represented as $t' = t^\star + (x_t, y_t, z_t)$ and optimized under the constraint of $x_t^2 \leq 0.1$ m, $y_t^2 \leq 0.1$ m, and $z_t^2 \leq 0.1$ m. Initially, $x_t = y_t = z_t = 0$ and

$t^\star$ is estimated as

$$t^\star = \frac{1}{M}\sum_{i=1}^{M} p_{a_i} - \frac{1}{M}\sum_{i=1}^{M} \hat{R} p_{b_i} \tag{4.8}$$

Therefore, $t^\star$ represents the distance between the centroid of both trajectories (analogous to [Aru87]). The optimization of t' under the given constraints yields a translational offset $\hat{t}'$ that cannot be too far away from the initially estimated translation.

As a result, $\hat{R}$ and $\hat{t}$, or rather $\hat{t}'$, yield the calibration matrix from *device B* to *device A*. Figure 4.3 shows the movement of two devices that were calibrated using the trajectory-based calibration algorithm.

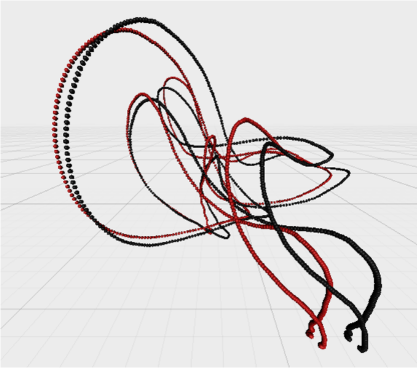


Figure 4.3: The calibrated trajectory of two interlinked devices that were picked up from a table, moved around the room, and then set back on the table again.

4.1.1 Evaluation

Setup To evaluate the performance of the trajectory-based calibration algorithm proposed in the previous section, two other algorithms commonly used for calibration and tracking were implemented for comparison.

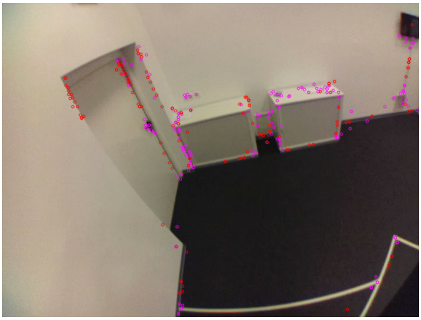


Figure 4.4: SIFT features within an image from the evaluation data set. Pink points depict successfully reconstructed features. Red points show features that could not be matched to the scene.

One algorithm is based on the structure from motion (SfM) technique [Koe91, Ull79]. SfM algorithms allow each device to generate a virtual map of a user's surroundings and track its position and rotation in this environment using camera and IMU input. After that, the virtual maps of two users can be aligned to get the calibration matrix. This approach is comparable to shared anchor methods[1]. COLMAP [Sch16a, Sch16b] is used for the SfM calibration. Its algorithm is supplied with a series of pictures from the cameras of both devices. COLMAP then finds SIFT features (scale-invariant feature transform) [Low04] in each image to obtain the relative camera poses for each image which in turn gives the calibration between the two devices (see Fig. 4.4).

The other algorithm uses marker detection to calibrate two devices. This approach requires each device to have a camera and one fiducial marker to be placed in the room. Systems such as Chilitags [Bon13] or AprilTags [Ols11, Wan16] can be used to detect the marker and calculate the position and orientation of the devices with respect to its location. This yields a shared calibration for both devices. Before the marker detection algorithm, the device's cameras were calibrated using Kalibr [Reh16][2].

The implemented system is designed as a service and features an easy-to-use

---

[1] https://developers.google.com/ar/develop/java/cloud-anchors/overview-android/
https://azure.microsoft.com/de-de/services/spatial-anchors/

[2] COLMAP finds the camera calibration matrix during the solving process.

GUI to allow quick calibration. The GUI is implemented in Unity as it allows to deploy the system to a diverse set of devices. All calculations are done on a server so that the system is also appropriate for hardware with limited resources, e.g., computing power or battery life, and is independent of an operating system. The calibration suite currently offers the three calibration algorithms described above but can be extended further. Two HTC Vive Pro HMDs were used as their Lighthouse tracking system is very precise if tracking is not lost [Nie17b]. To assure maximum accuracy, the SteamVR calibration setup for each Vive Pro was executed before each test and in case the tracking was lost. The Lighthouse system tracks one or more VR devices using infrared light emitted from two base stations which is captured by several diodes on the devices.

To evaluate the calibration accuracy of the three algorithms, a ground truth calibration was calculated. For this, the positions of the Lighthouse base stations for both Vive Pro HMDs were retrieved. By matching the position of both base stations, the calibration matrix from one Vive to the second Vive can be calculated.

Tasks and Procedures 30 seconds of data were recorded for each of the following calibration setups:

I. Trajectory-based calibration algorithm:

I.1 Slow translation
I.2 Fast translation
I.3 Slow rotation
I.4 Fast rotation
I.5 Slow spherical movement
I.6 Fast spherical movement
I.7 Slow spontaneous movement (*)
I.8 Fast spontaneous movement (*)

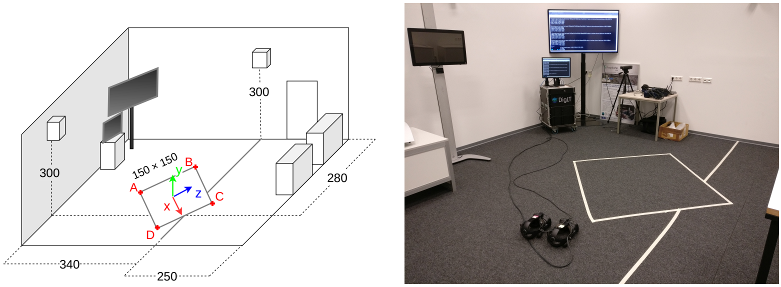
II. SfM-based calibration algorithm:

II.9 Table scene (*)
II.10 Room corner scene (*)

III. Marker-based calibration algorithm:

III.11 Moving camera (*)
III.12 Static camera

Some movements are predefined and could be executed without inviting other users. However, all data recordings highlighted with (*) were executed by users because their movement cannot be defined or is open to interpretation. Each data set contains N=44 records. For the recording of the user data sets, 11 participants were invited to perform the tasks. In the case of a user data set, each of the 11 users performed 4 calibration movements. All positional data was recorded at 60 Hz. *Device A's* camera recorded videos with 40 Hz and *device B* had a camera sampling rate of 30 Hz. All trajectory-based movements (I) were executed in two ways, i.e., slow and fast movements, to evaluate the calibration difference at different speeds. For I.1 – I.6, the movements were executed with a factor of f_m=3 in the slow condition and f_m=6 in the fast condition. All movements were executed inside the calibration space defined by the square *ABCD* (see Fig. 4.5a and 4.5b).



(a) Room setup measurements.

(b) Movement area (I.1 – I.8).

Figure 4.5: Setup for the room and Trajectory-based algorithm.

For translational movements, I.1 and I.2, the two HMDs that form the rigid body system are oriented to face the negative x-direction and positioned in the center of the calibration space. Then, the following translational movements were executed along the different axis:

$$([-x \leftarrow][x \rightarrow][x \rightarrow][-x \leftarrow]) \times f_m$$
$$([-z \leftarrow][z \rightarrow][z \rightarrow][-z \leftarrow]) \times f_m$$
$$([-y \leftarrow][y \rightarrow][y \rightarrow][-y \leftarrow]) \times f_m$$

For rotational movements, I.3 and I.4, the two HMDs that form the rigid body system are oriented to face the negative x-direction and positioned in the center of the calibration space. Then, the following rotational movements were executed along the different axis:

$$([-y \circlearrowleft][y \circlearrowright][y \circlearrowright][-y \circlearrowleft]) \times f_m$$
$$([z \circlearrowright][-z \circlearrowleft][-z \circlearrowleft][z \circlearrowright]) \times f_m$$
$$([-x \circlearrowleft][x \circlearrowright][x \circlearrowright][-x \circlearrowleft]) \times f_m$$

For spherical movements, I.5 and I.6, the two HMDs that form the rigid body system are oriented to face the negative x-direction. The system is moved 75 cm in negative x-direction in between points *A* and *B* (see Fig. 4.5a). Then, the following spherical movements were executed by rotating the HMDs along the different axis around the center of the calibration space:

$$([-y \curvearrowleft][y \curvearrowright][y \curvearrowright][-y \curvearrowleft]) \times f_m$$
$$([z \curvearrowright][-z \curvearrowleft][-z \curvearrowleft][z \curvearrowright]) \times f_m$$
$$([-x \circlearrowleft][x \circlearrowright][x \circlearrowright][-x \circlearrowleft]) \times f_m$$

For spontaneous movements, I.7 and I.8, the participants were informed about the goal of their movement. They were asked to perform any movement that comes to their mind while moving ‘not too fast’ in the slow condition and ‘move fast’ in the fast condition. Users were encouraged to find a suitable grip that fixes the relative position of the headsets. Also, users were asked not to leave the calibration space during the movements.

Two scenes were prepared for the SfM algorithm (II). The video stream was displayed on a large display so that users could judge the camera’s viewing angle. Users first captured the scenes using *device A*. Then each user performed four calibrations using *device B*. Users were asked to ‘not move too fast’ to avoid camera blur. Users were encouraged to capture the scenes from different angles and at different distances.

A table with a prepared scene (II.9) was placed so that the leading edge of the table was parallel to the line *AB* (see Fig. 4.5a). The scene was filled with objects that contained many color contrasts and brightness variations, to obtain as many SIFT features as possible (see Fig. 4.6a).

The room corner scene (II.10) contains mostly large and white surfaces (see Fig. 4.6b). Users were asked to not get closer towards the target as 2.5 m, which was indicated by a line on the ground (see Fig. 4.5a). This data recording, therefore, contains an environment with large monotonous surfaces but movements with large spatial extension.

(a) SfM table scene (II.9). **(b)** SfM room corner (II.10).

Figure 4.6: Setup for SfM-based algorithm.

The marker-based calibration (III) used the same table that was placed in the room as described for II.9. An 8×6 AprilTag board, printed on an A3 sheet, was placed on the table (see Fig. 4.7). The marker board was placed orthogonal to the table, with an 8° rotational offset towards the ceiling of the room. Again, users first captured the scene using *device A* and then performed four calibrations using *device B*.

For the moving camera (III.11), users were asked to 'not move too fast' to avoid camera blur, to perform movements within a distance of 20 – 60 cm, and to avoid flat angles to keep the perspective distortion of the markers at a minimum.

For the static camera (III.12), the headset was placed on a box facing the marker board at a distance of 30 cm and a height of 20 cm. The headset was slightly facing up so that the integrated cameras directly faced the center of the markers.

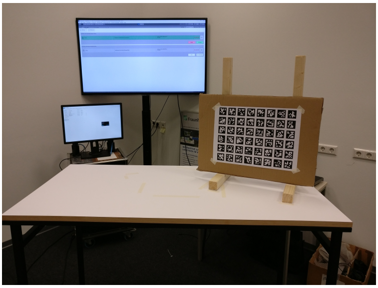


Figure 4.7: Setup for Marker-based algorithm (III.11 and III.12).

Measurements The three algorithms are executed under different conditions, i.e., changes in movement or room setup. These varying conditions might impact the precision of the calibration result. Furthermore, the algorithms may be differently suited for the task of calibrating two devices. Because of that, the following hypotheses were formulated for the results of the evaluation:

H_{A_X} for $X =$ algorithm I, II, and III. For each of the algorithms, there exists at least one condition that achieves a significantly better calibration than the others.

H_B One algorithm (I, II, or III) reaches a significantly better calibration than the others.

The rotation $R_A = \Delta R \cdot R_{GT}$ calculated by an algorithm A (i.e., I, II, or III) differs from the reference rotation R_{GT} by an error ΔR. ΔR can be represented as an angle and an axis. The angle describes the rotational error and can be calculated as $\theta_{\Delta R} = 2 \cdot arccos(|q_A \cdot {q_{GT}}^{-1}|)$ with q_A and q_{GT} as the unit quaternions of R_A and R_{GT} [Jai19]. The translational error is described as the euclidean distance between the position vectors contained in the calibration and ground truth matrices. To remove outliers, points with a larger distance than 3 standard deviations from the average value were excluded from the results.

4.1.2 Results

Table 4.1: Statistical values for the translational and rotational errors of the cleaned data sets.

	Translational error (in m)				Rotational error (in °)			
Data set	N	M	SD	95% CI	N	M	SD	95% CI
[I.1]	44	0.1056	0.0154	0.1010	43	1.2288	0.4389	1.0961
[I.2]	44	0.1177	0.0094	0.1149	44	1.9148	0.8080	1.6733
[I.3]	44	0.0196	0.0073	0.0174	44	0.7474	0.2982	0.6583
[I.4]	44	0.0237	0.0039	0.0225	43	0.9718	0.1603	0.9233
[I.5]	42	0.0290	0.0115	0.0255	43	0.6000	0.2430	0.5265
[I.6]	44	0.0317	0.0042	0.0304	44	1.0746	0.1747	1.0224
[I.7]	42	0.0595	0.0456	0.0455	42	1.6462	1.3073	1.2461
[I.8]	43	0.0480	0.0287	0.0393	43	1.4547	1.1148	1.1175
[II.9]	44	0.0904	0.0540	0.0742	44	2.9294	2.3461	2.2282
[II.10]	44	0.0590	0.0379	0.0477	44	1.4075	0.8107	1.1651
[III.11]	44	0.5138	0.3034	0.4231	44	20.0404	16.7612	15.0306
[III.12]	44	0.1505	0.0673	0.1304	44	4.5872	2.0914	3.9621

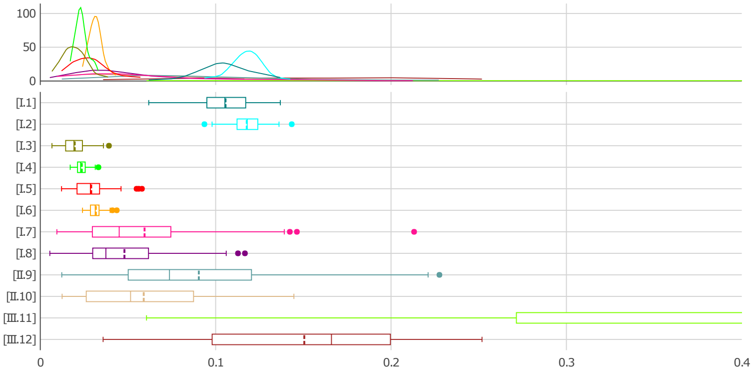


Figure 4.8: Distribution of the translational error (in m) of the cleaned data sets. The x-axis is truncated to increase the resolution for the lower errors. Significant differences are not indicated.

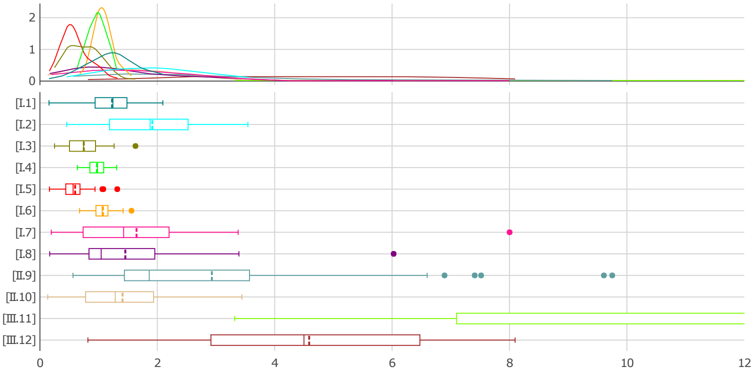


Figure 4.9: Distribution of the rotational error (in °) of the cleaned data sets. The x-axis is truncated to increase the resolution for the lower errors. Significant differences are not indicated.

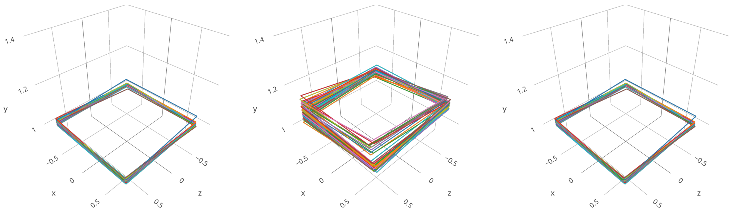
Quantitative Results The translational and rotational errors are described in Table 4.1. The distributions of the errors are displayed in Figures 4.8 and 4.9. A repeated-measures ANOVA for the translational error of the eight trajectory-based calibration algorithms (I) yields a significant difference (F(7,339)=148.600, p<.001, N=347) with a large effect size (partial η^2=.754). A Tukey HSD test yields the following: The data sets form three clusters where each data set is not significantly different from the other data sets in the cluster but is significantly different from any data sets from the other clusters. All significant differences have a value of p<.001. The three clusters are slow and fast translation (I.1 and I.2), slow and fast spontaneous movement (I.7 and I.8), and slow and fast rotation and spherical movement (I.3 – I.6). A repeated-measures ANOVA for the rotational error also yields a significant difference (F(7,338)=17.650, p<.001, N=346) with a large effect size (partial η^2=.268). A Tukey HSD test shows that the data set with the lowest error, I.5, has no significant differences to the data sets I.3, I.4, and I.6. However, I.5 is significantly different from all other data sets with p<.001. I.2 is significantly different from I.1, I.3, I.4, and I.6 with p<.001. I.7 is significantly different from I.3 and I.4 with p<.001, and from I.6 with p=.005. Also, I.8 is significantly different from I.3 with p<.001.

A repeated-measures ANOVA for the translational error of the two SfM-based calibration algorithms (II) yields a significant difference between II.9 and II.10 (F(1,86)=9.970, p=.002, N=88) with a medium effect size (partial η^2=.104). A repeated-measures ANOVA for the rotational error yields a significant difference between II.9 and II.10 (F(1,86)=16.54, p<.001, N=88) with a large effect size (partial η^2=.161).

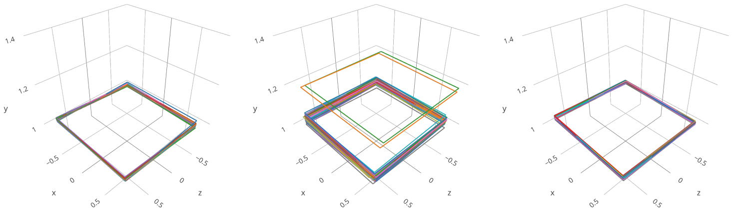
A repeated-measures ANOVA for the translational error of the two Marker-based calibration algorithms (III) shows a significant difference between III.11 and III.12 (F(1,86)=60.130, p<.001, N=88) with a large effect size (partial η^2=.411). For the rotational error, a repeated-measures ANOVA also measures a significant difference between III.11 and III.12 (F(1,86)=36.830, p<.001, N=88) with a large effect size (partial η^2=.300).

The hypothesis H_B is checked based on the best data sets from the three algorithms. The test includes I.3 and I.5, as well as II.10 and III.12. A repeated-measures ANOVA for the translational error shows a significant difference between the four conditions (F(3,170)=100.400, p<.001, N=174) with a large effect size (partial η^2=.639). A Tukey HSD test yields the following: Again, I.3 and I.5 are not significantly different. However, all other differences are significant with p<.001, except for I.5 and II.10 with p=.003. A repeated-measures ANOVA for the rotational error yields a significant difference (F(3,171)=117.500, p<.001, N=175) with a large effect size (partial η^2=.639). A Tukey HSD test yields the following: As for the translation, I.3 and I.5 are not significantly different. Algorithm III with III.12 is significantly worse than all other algorithms with p<.001. Algorithm I (I.5) and II (II.10) have a significant difference with p=.006.

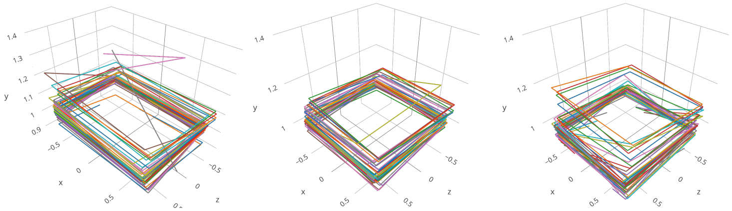
Qualitative Results To evaluate the quality of the different calibration results, it was simulated how the translational and rotational errors would impact the user's experience during usage. For this, a *device B* was positioned at the center of the simulated space at a height of 1 m. *Device A* is moved around the XZ-plane in the shape of a square with an edge length of 1.5 m at a height of 1 m. For each algorithm and each data set, the ground truth matrix T_{GT} and the calculated calibration matrix $T_C = \Delta T \cdot T_{GT}$ with some calibration error



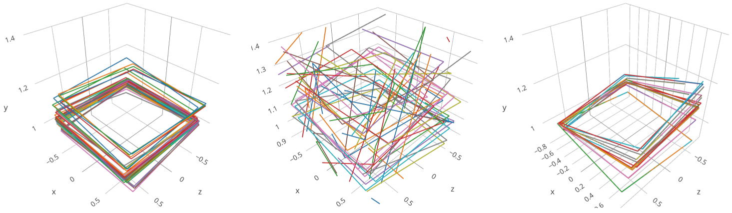
(a) [I.1] Slow translation. **(b)** [I.2] Fast translation. **(c)** [I.3] Slow rotation.



(d) [I.4] Fast rotation. **(e)** [I.5] Slow spherical movement. **(f)** [I.6] Fast spherical movement.



(g) [I.7] Slow spontaneous movement. **(h)** [I.8] Fast spontaneous movement. **(i)** [II.9] Table scene.



(j) [II.10] Room corner. **(k)** [III.11] Moving camera. **(l)** [III.12] Static camera.

Figure 4.10: Simulated trajectories that show the error introduced during the calibration process for a device that was moved along the sides of a square.

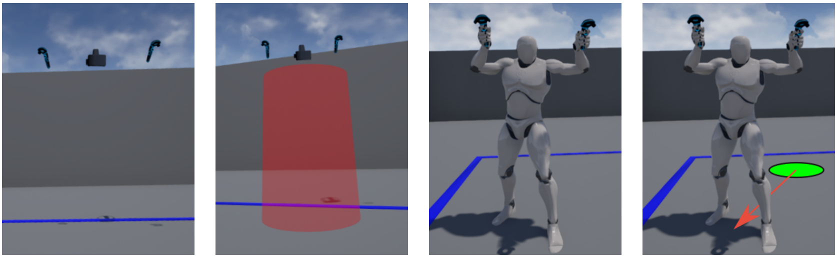
ΔT were recorded. The movement of *device A* is now projected into an uncalibrated space by transforming it using the inverse ground truth matrix and then reprojected into calibrated space by transforming it using the calibration matrix. The resulting point $p_i' = T \cdot T_{GT}^{-1} \cdot p_i = \Delta T \cdot T_{GT} \cdot T_{GT}^{-1} \cdot p_i = \Delta T \cdot p_i$ contains the error that is introduced by the calibration process. The series (p_i') of the simulated points for each algorithm and data set are shown in Figure 4.10. It highlights the strong effects of higher variances, larger rotational errors, and outliers.

4.1.3 Discussion

Overall, H_{A_I}, $H_{A_{II}}$, $H_{A_{III}}$, and H_B can be supported. As the results of the evaluation and the simulated trajectories show, algorithm I provides a reliable and accurate way to calibrate two devices towards a shared coordinate system. Slow rotational or spherical movements provide better results than fast movements or translational and spontaneous movements. Furthermore, slow rotations can be executed easily in a small space without much effort. However, maintaining a constant offset between the devices seems to be crucial to reduce errors.
The SfM algorithm achieves better calibration results in the room corner scene. This was not expected, as the table scene should provide more SIFT features. However, the larger movements and different viewing angles seem to result in a more precise calibration.
The marker-based algorithm seems unsuitable for VR headsets with a wide FOV camera. This supports the work of Zhen et al. [Zhe17] who show that a Vive camera can yield errors of several cm at short distances.

4.2 Collision Avoidance for Co-Located Collaboration

After users have calibrated their VR devices towards a shared coordinate system, they can explore the virtual worlds collaboratively. But, as VR HMDs



(a) Basic avatar (BA) **(b)** Cylinder avatar (CA) **(c)** Fullbody avatar (FA) **(d)** Shifted avatar (SA)

Figure 4.11: The four different avatars that are compared regarding their capability to avoid collisions. The green circle of the shifted avatar represents the real position of the user and the avatar is shifted towards the location of the current user.

block out the real world, users cannot see each other anymore. This can lead to users running into each other and hurting themselves. A user avatar might allow users to estimate the location of others and therefore to avoid collisions. However, different avatar representations can be used. Several VR applications use a very basic user representation and only show the HMD and the two controllers that the user holds, i.e., the basic avatar (BA) (see Fig. 4.11). This representation is easy to implement since the location of the hardware is known and no additional effort is needed to use the BA. However, the floating HMD and controllers are very subtle. To increase the visibility of this avatar, it was fitted with a bounding cylinder that is centered around the user's head and can be easily seen by other users. Due to its strong indication, the cylinder avatar (CA) highlights the current location of others and aids users' perception. With the use of pose tracking, a full-body avatar (FA) can be displayed. The FA brings all the movements of the user into VR. If full-body tracking is not available or unwanted it is furthermore possible to use inverse kinematics to estimate the location of the joints from the pose of the HMD and controllers. Since the goal was to minimize effort and hardware cost to integrate the avatars in any application, the second option was chosen. Because only the information about the head and hands was available, the legs of the FA did not move. If additional tracking is added to the feet this issue can be resolved. Because of network or application lag, it is possible that during a quick motion a user avatar is not updated as quickly as it needs to. Also, users tend to underestimate distances while using an HMD. The result might be that another user appears further away than she/he really is. With the shifted avatar (SA), a user is visualized 0.5 m closer in VR than she/he is in the real world, thereby introducing a buffer to further decrease the risk of a collision.

4.2.1 Evaluation

Setup To evaluate the capabilities of the different avatars, a scenario was designed that would provoke collisions. In a competitive task, two users were asked to reach a target object as fast as possible without colliding with the other participant. The objects were placed at two different distances (evenly distributed) and so that the paths of the users crossed (see Fig. 4.12). The application was implemented using Unreal Engine 4 and used two HTC Vive HMDs to present the virtual content to the users. Since the HMD cables of the users could entangle and lead to users falling and a wireless connection for the HMDs was not available, the cables to the HMDs were suspended from the ceiling of the room. The length of the cable was adjusted with a retractable dog leash.

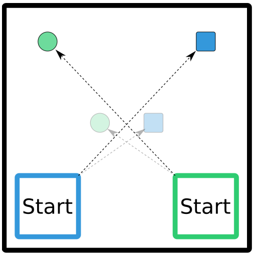


Figure 4.12: The setup of the competitive user study. Both participants need to access an object but their paths' cross. The objects were placed at different distances with an even distribution.

Participants The four presented avatars were compared in a 28 participant within-subject user study. The participants (19 male, 9 female) were M=22.5 (SD=2.6) years old. On a five-point Likert scale from 0 (none) to 4 (high) the participants rated their experience with computer games as M=3.286 (SD=0.810) and with VR as M=1.107 (SD=1.100).

Tasks and Procedures For each avatar, a set of eight rounds, i.e., three training rounds and five timed rounds, was performed. A round begins with a countdown and both users standing in their respective start zone. The round ends when a user grabs the object assigned to her/him. At the end of a set, users were presented with a questionnaire containing six questions rated on a five-point Likert scale. Two participants performed a set using the same avatar. Each pair used all four avatars and performed four sets in total. The order of the used avatars was randomized to compensate for training and fatigue effects. At the very end, a final questionnaire was given to the participants.

Measurements The following hypothesis was formulated for the results of the evaluation:

H_A The different avatar appearances have a significant impact on users' capability to avoid collisions with better results for more extensive representations.

The speed of users was recorded during the tasks with the various avatars. In addition to that, a custom questionnaire assessed the safety, usability, and preference of the avatar representations. Users rated different questions on a five-point Likert scale from 0 (best) to 4 (worst). For the number of collisions, users rated the avatars with 0 (no collisions), 1 (at least one collision), 2 (at least three collisions), 3 (at least five collisions), and 4 (at least ten collisions).

4.2.2 Results

Quantitative Results The duration of each round was measured (see Fig. 4.13). Users took M=2.827 seconds (SD=1.797) to complete the task

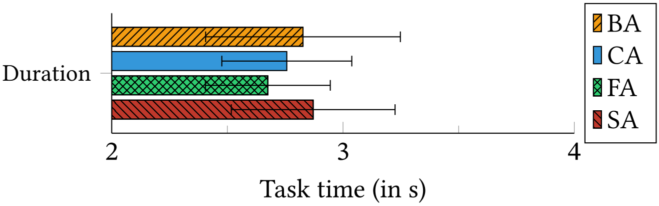


Figure 4.13: Time until one user reached the target. Results as average value with 95% confidence intervals.

with BA, M=2.757 seconds (SD=1.200) with CA, M=2.675 seconds (SD=1.151) with FA, and M=2.871 seconds (SD=1.512) with SA. Mauchly's test indicates a violation of sphericity (χ^2(5)=44.004, p<.001, Greenhouse-Geisser ϵ=.680). A Greenhouse-Geisser-corrected repeated-measures ANOVA for the duration yields no significant difference between the conditions (F(2.041,140.807)=0.401, p=.675, N=70).

Qualitative Results The results of the custom questionnaire (see Fig. 4.14) show that all avatars are suitable to predict the user's own location. Users rate the avatars as M=0.714 (SD=0.897) for BA, M=0.929 (SD=0.813) for CA, M=0.607 (SD=0.567) for FA, and M=1.143 (SD=1.008) for SA. A Friedman test shows no significant differences for the predictability of user's own location (χ^2(3)=6.910, p=.075, N=28). The avatars allow predicting the position of the other co-present user with M=1.643 (SD=1.254) for BA, M=1.286 (SD=0.937) for CA, M=0.821 (SD=0.905) for FA, and M=2.071 (SD=1.052) for SA. A Friedman test shows significant differences for the conditions (χ^2(3)=18.448, p<.001, N=28) with a small effect size (Kendall's W=.220). Pairwise Bonferroni-corrected tests show a significant difference between FA and SA (p=.001). Users have a medium-low effort to avoid collisions with each other. They rate the avatars as M=1.893 (SD=1.133) for BA, M=1.143 (SD=1.008) for CA, M=1.536 (SD=0.999) for FA, and M=1.893 (SD=1.066) for SA. A Friedman test shows significant differences for the conditions (χ^2(3)=10.643, p=.014, N=28) with a small effect size (Kendall's W=.127). However, pairwise Bonferroni-corrected tests show no significant differences between the avatars. All avatars emit a medium to

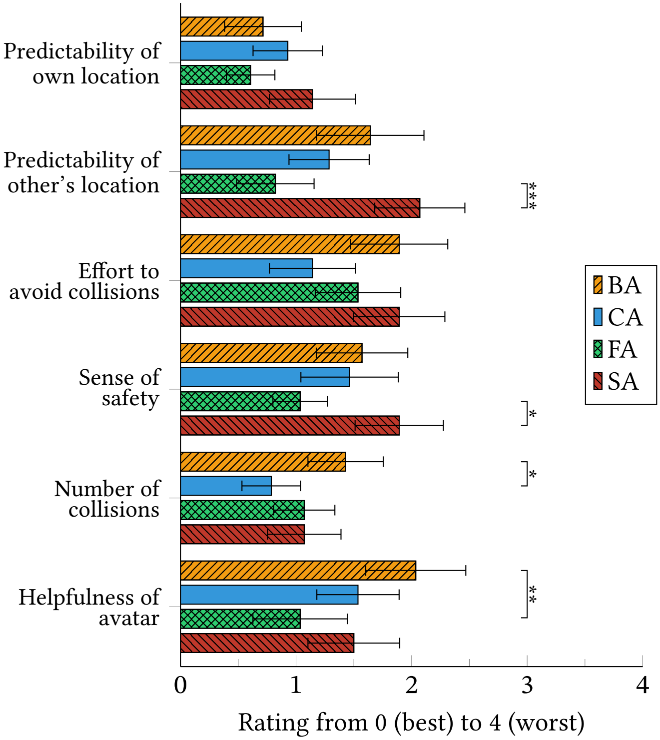


Figure 4.14: Custom questionnaire. Results as average value with 95% confidence intervals and asterisks indicating significant differences.

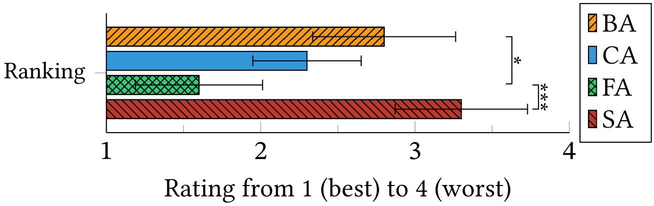


Figure 4.15: Preference rating of users that noticed the shift. Results as average value with 95% confidence intervals and asterisks indicating significant differences.

high sense of safety with M=1.571 (SD=1.069) for BA, M=1.464 (SD=1.138) for CA, M=1.036 (SD=0.637) for FA, and M=1.893 (SD=1.031) for SA. A Friedman test shows significant differences for the conditions (χ^2(3)=11.589, p=.009, N=28) with a small effect size (Kendall's W=.138). Pairwise Bonferroni-corrected tests show a significant difference between FA and SA (p=.027). Users experienced some collisions with all avatars with M=1.429 (SD=0.879) for BA, M=0.786 (SD=0.686) for CA, M=1.071 (SD=0.716) for FA, and M=1.071 (SD=0.858) for SA. A Friedman test shows significant differences for the conditions (χ^2(3)=11.441, p=.010, N=28) with a small effect size (Kendall's W=.136). Pairwise Bonferroni-corrected tests show a significant difference between BA and CA (p=.037). Lastly, users rated the helpfulness of the different avatars to avoid collisions with each other. The avatars achieved M=2.036 (SD=1.170) for BA, M=1.536 (SD=0.962) for CA, M=1.036 (SD=1.105) for FA, and M=1.500 (SD=1.072) for SA. A Friedman test shows significant differences for the conditions (χ^2(3)=15.448, p=.001, N=28) with a small effect size (Kendall's W=.184). Pairwise Bonferroni-corrected tests show a significant difference between BA and FA (p=.006).

To assess the effects of the shift of the SA, the participants were not told that the avatar of the other party was shifted towards them. 20 users (71.4%) noticed the shift of the fourth avatar and felt unsafe since they could not estimate the actual location of the other user. Those users ranked the four avatars as shown in Fig. 4.15. The ratings yield M=2.800 (SD=1.056) for BA, M=2.300 (SD=0.801) for CA, M=1.600 (SD=0.940) for FA, and M=3.300 (SD=0.979) for SA. A Friedman test shows significant differences for the avatars (χ^2(3)=18.960, p<.001, N=20) with a medium effect size (Kendall's W=.316). Pairwise Bonferroni-corrected tests show a significant difference between FA and BA (p=.020), and FA and SA (p<.001).

4.2.3 Discussion

The results of the user study show, that the representation of users' avatars influences the number of collisions and the user's sense of safety. Therefore H_A can be supported.

The minimal representation of the BA is not very suitable for local multi-user interaction, as it has the highest amount of collisions in the user study. In addition to that, it is rated worse than the other avatars in many categories and is never rated better in any category.

The CA shares the lowest amount of collisions with the FA[3]. Furthermore, the effort to avoid collisions is low. Its representation is very easy to implement and does not need any additional calculations for avatar animation or additional hardware setup. It can be integrated into an existing system very easily. However, the strong indication of the user might disturb the immersion of the VR experience.

The FA leads to a high sense of safety and a low number of collisions and is preferred by the participants of the user study. Additionally, it facilitates the predictability of the other user's location. The representation of the avatar needs body pose tracking or inverse kinematic calculations. For example, inside-out pose tracking can be achieved with an additional head-worn camera [Rho16]. The implemented inverse kinematic calculations are only an estimate of the user's body pose and contain a high error for the legs. A simple solution to this problem is a floating avatar without legs[4]. Another way to increase the accuracy of this method is to attach additional trackers to the user's feet [Kim13, Mou17], wrists [Kim13, Ria15], or hip [Ria15]. Also, gait can be reconstructed via analyzing the movement of the HMD [Cas16, Fei20].

The shifted version of the full-body avatar leads to bad predictability of the location of the other user since it is intentionally adjusted. Moreover, this modification leads to bad ratings and insecurity among users. The shift is readily noticed and should be avoided.

The presented results are in-line with the related work [Pod17, Sca17, Wil19].

---

[3] It should be noted that the given task was designed to provoke collisions and it can be expected that a small amount of collisions always occurs.

[4] https://www.oculus.com/facebook-horizon/

4.3 Comparison of Co-Located and Remote Collaboration

By blocking out the real world, VR HMDs can visualize immersive 3D content through a stereoscopic presentation. With the addition of a headset, 3D audio can be presented as well. To maintain communication with other co-located users, direct speech may be mixed with virtual acoustic feedback. However, as users start exploring the virtual worlds independently from one another, digital speech communication is necessary to represent the other users' voices from their respective virtual avatar locations. As the visual and auditory senses of the users are now overwritten by the VR system, the question arises whether the other user needs to be physically present since she/he is not visible nor audible at all. Without any changes towards the application, the other user could also join the virtual room via a network connection from another location. Therefore, a task to compare co-located and remote collaboration in VR was designed.

4.3.1 Evaluation

Setup As with the collision avoidance evaluation, two HTC Vives with suspended cables were used for the co-located scenario. For the remote scenario, one additional HTC Vive was set up in a separate room, however without a suspended cable. The two rooms are located in different buildings but connected by a 1 Gbps Ethernet network. In both scenarios, users used a Logitech G930 headset to communicate via voice chat with each other. The direction of the audio signal of a speaking user is adjusted according to the location of her/his avatar.

A user is represented by an avatar that is aligned using the head and hand positions through inverse kinematics (see Fig. 4.16). The avatar is important for the feeling of co-presence [Ste15]. The stylized representation is not significantly different from a human avatar [Ger01] and avoids the uncanny valley effect [Mor12]. Roth et al. [Rot16] determined that nonrealistic avatars handicap social interactions. However missing behavioral characteristics, like gaze or facial expressions, can be partially compensated by using other behavior channels, like gestures. They concluded that a mannequin is a universal representation of a human which is easy to reproduce and animate.

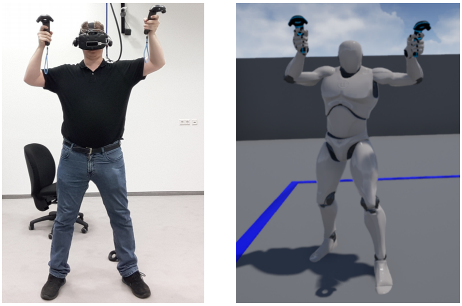


Figure 4.16: Avatar representation of the user in VR using inverse kinematics.

Participants A within-subject user study with 28 participants who performed tasks in pairs of two was conducted. The participants (19 male, 9 female) were M=22.5 (SD=2.6) years old. On a five-point Likert scale from 0 (none) to 4 (high) the participants rated their experience with computer games as M=3.286 (SD=0.810) and with VR as M=1.107 (SD=1.100).

Tasks and Procedures To assure user collaboration, a knowledge-transfer scenario was implemented where two users take different roles. One user, the *expert*, highlights virtual objects for another user using a pointing gesture. The second user, the *worker*, then needs to interact with the indicated object by selecting it using direct touch. To evaluate the effects of the two collaboration setups, three different pointing gestures were examined. The used pointing gestures are virtual hand, virtual pointer [Pou98a], and target marker (see Fig. 4.17). With target marker, the expert has a virtual laser pointer attached to her/his hand. In addition to that, the pointed-at object is highlighted. The worker also sees the visual highlight but not the beam of the laser pointer. The virtual objects are represented by cubes, arranged in a grid of 3×3×3 as in [Win02]. To increase task difficulty the grid has a static and a rotating mode. In the rotating mode, the whole grid rotates slowly around two axes with different velocities.

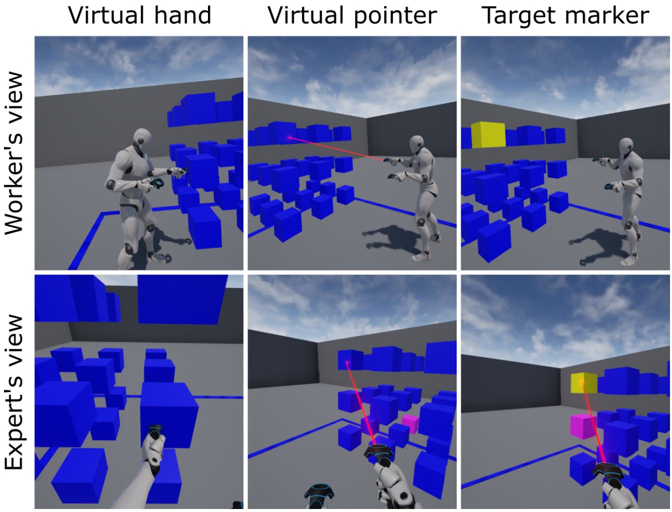


Figure 4.17: All three pointing techniques from left to right with the worker's view on top and the expert's view on the bottom.

Each pair performed the tasks in both the remote scenario as well as the co-located scenario. The roles of the users were switched when the scenario changed. Both users performed all three gestures in the role of the expert. To minimize the effects of learning and fatigue on the results, the order of the scenarios, roles, and gestures was randomized. A task consists of three training rounds (two with static grid, one with rotating grid) and six timed rounds (three with static grid, three with rotating grid). One round contained one indication of the expert and the interaction of the worker with the virtual object. The round starts with both users standing on designated start positions and ends with the selection of the virtual object by the worker.

Measurements Based on the results of related work, the latency-free mediated collaboration should be independent of users' physical location. Because of that, the following hypotheses were formulated for the results of the evaluation:

H_A The task execution using the different gestures is not significantly different while working co-located and remotely.

H_B The perceived co-presence of users is not different in both setups.

The time users took to complete the tasks was measured. Furthermore, subjective results were recorded using custom measurements, such as preference, feeling of co-presence, collaboration, and nausea on a five-point Likert scale.

4.3.2 Results

Quantitative Results Average interaction times per round are about four to five seconds, as shown in Fig. 4.18. In the co-located setup, users take M=4.815 seconds (SD=1.986) with virtual hand, M=3.926 seconds (SD=1.925) with virtual pointer, and M=3.816 seconds (SD=1.540) with target marker. The remote setup is slightly slower with M=5.099 seconds (SD=2.008) for virtual hand, M=4.422 seconds (SD=1.823) for virtual pointer, and M=4.786 seconds (SD=3.036) for target marker. Given the continuous completion time and dependent measurements, a two-way repeated-measures ANOVA with the location (i.e., co-located and remote) and pointing technique (i.e., virtual hand, virtual pointer, and target marker) as factors was used. Mauchly's test is not calculated for the location as sphericity is always met for two levels. Mauchly's test indicates a violation of sphericity for the pointing technique ($\chi^2(2)$=8.342, p=.015, Huynh-Feldt ϵ=.931). No violation of the assumption of sphericity is indicated for location*technique ($\chi^2(2)$=0.435, p=.805). There is a significant difference for location (F(1,83)=9.710, p=.003, N=84) with a medium effect size (partial η^2=.105) and pointing technique (F(1.862,154.579)=9.150, p<.001, N=84) with a medium effect size (partial η^2=.099). Pairwise Bonferroni-corrected tests show that virtual hand is significantly slower than virtual pointer (p<.001) and target marker (p=.001). There are no significant differences for location*pointing technique (F(2,166)=2.110, p=.125, N=84).

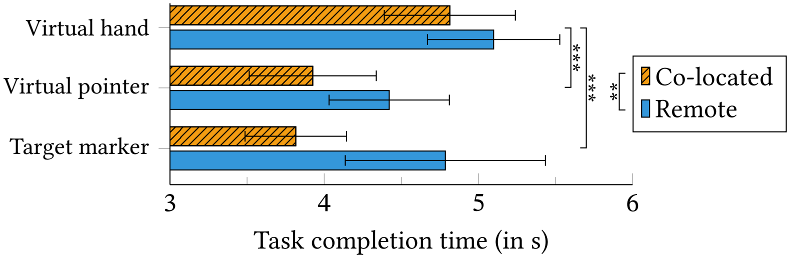


Figure 4.18: Task completion time by pointing technique and location. Results as average value with 95% confidence intervals and asterisks indicating significant differences.

Since the tasks are identical in both setups, further investigations were performed. The time it took the expert to point to the target object for the first time was recorded (see Fig. 4.19). In the co-located setup, users take M=1.366 seconds (SD=1.325) with virtual hand, M=2.748 seconds (SD=1.430) with virtual pointer, and M=2.812 seconds (SD=1.371) with target marker. The remote setup is slightly slower with M=1.450 seconds (SD=0.993) for virtual hand, M=3.182 seconds (SD=1.393) for virtual pointer, and M=3.391 seconds (SD=2.147) for target marker. Again, Mauchly's test is not calculated for the location as sphericity is met. Mauchly's test indicates a violation of sphericity for the pointing technique (χ^2(2)=20.288, p<.001, Huynh-Feldt ϵ=.835). No violation of the assumption of sphericity is indicated for location*technique (χ^2(2)=1.287, p=.525). There is a significant difference for location (F(1,83)=9.518, p=.003, N=84) with a medium effect size (partial η^2=.103) and pointing technique (F(1.669,138.535)=82.232, p<.001, N=84) with a large effect size (partial η^2=.498). Pairwise Bonferroni-corrected tests show that virtual hand is significantly faster than virtual pointer and target marker (both p<.001). There are no significant differences for location*pointing technique (F(2,166)=1.727, p=.181, N=84).

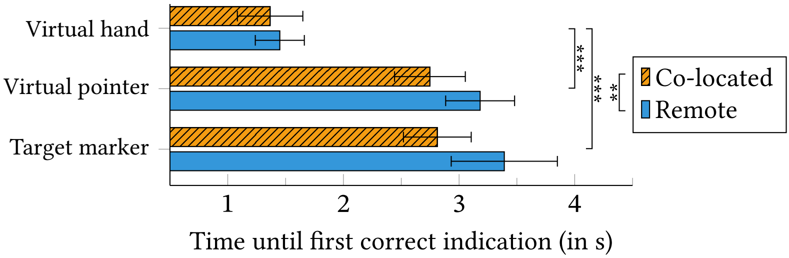


Figure 4.19: Time until the expert indicated the correct target for the first time by pointing technique and location. Results as average value with 95% confidence intervals and asterisks indicating significant differences.

Qualitative Results Figure 4.20 shows the experts' preferences for the different pointing gestures sorted by the room setup[5]. Participants were asked to rank the gestures from 1 (first place) to 3 (last place). In the co-located setup, experts rated the gestures as M=2.214 (SD=0.802) for virtual hand, M=2.214 (SD=0.802) for virtual pointer, and M=1.571 (SD=0.756) for target marker. While working remotely, experts rated the gestures as M=2.500 (SD=0.760) for virtual hand, M=1.857 (SD=0.535) for virtual pointer, and M=1.643 (SD=0.929) for target marker. Since no user performed the gestures in the role of the expert in both co-located and remote setup, the samples are independent and the Mann-Whitney-U-Test is used to check for significant differences. The differences between the location of users are not significant for the virtual hand method (U=77.500, p=.352, N=14), nor for the virtual pointer (U=70.500, p=.210, N=14), or target marker (U=98.000, p=1.000, N=14).

Users were asked how much they depended on speech communication while solving the tasks on a scale from 0 (not at all) to 4 (very much) (see Fig. 4.21). A WSR test for the ratings of M=0.464 (SD=0.637) in the co-located setup and M=0.393 (SD=0.497) in the remote setup shows no significant difference between the two scenarios (z=-0.577, p=.564, N=28). Furthermore, users rated

---

[5] Not shown for the worker as her/his selection gesture was always the same.

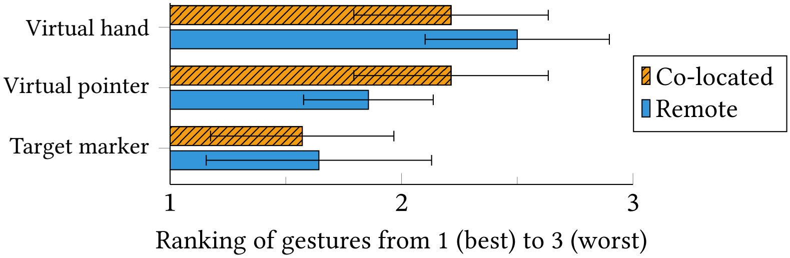


Figure 4.20: Ranking of the expert for the pointing gestures. Results as average value with 95% confidence intervals.

the amount of co-presence they experienced with the other user while performing the tasks of the user study. Co-presence was assessed on a scale from 0 (users feel like they are in different rooms) to 4 (users feel like they are in the same room). In the co-located setup users rated co-presence as M=3.179 (SD=0.905) and as M=2.821 (SD=0.905) in the remote setup. A WSR test shows that the difference is not significant (z=-1.696, p=.090, N=28). Participants were asked how pleasant the collaboration with the partner was. Users rated the collaboration on a scale from 0 (very unpleasant) to 4 (very pleasant) with M=3.929 (SD=0.262). This value shows that the pairs could work well together and the results are not negatively affected by a user's refusal to cooperate.

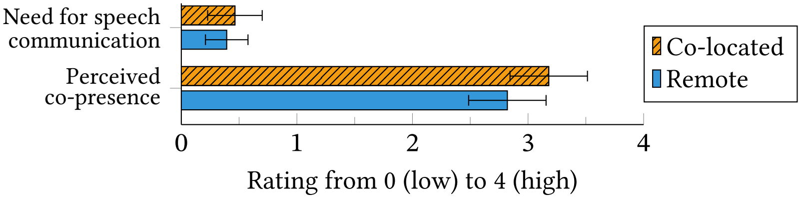


Figure 4.21: Requirement for speech communication and perceived co-presence. Results as average value with 95% confidence intervals.

When asked, 50.0% of the users did not prefer either one setup. 10.7% preferred the co-located interaction and 39.3% liked the remote interaction better. Ten out of eleven users explained their preference for the remote setup by saying that they did not need to worry about any collisions with the other party while working remotely. The remaining user was impressed by the capabilities of the collaboration via a network. From the three users who preferred the co-located setup, two said the collaboration is more realistic and one said that she/he did perceive the other user more as a human rather than a robot.

Additionally, users were asked if they experienced nausea to check if the collected data could be negatively influenced. Users reported almost no nausea with M=3.786 (SD=0.418) on a scale from 0 (strong nausea) to 4 (no nausea). The authenticity of the virtual world was experienced as M=2.607 (SD=0.737) on a scale from 0 (not real) to 4 (very real). Lastly, users were asked how much they were distracted by disturbances from outside the VR world. Users reported medium to low disturbance with M=2.786 (SD=0.957) on a scale from 0 (strongly) to 4 (not at all).

4.3.3 Discussion

H_A can only be partially supported. A comparison of the two scenarios, co-located and remote, shows no significant differences in user rating. But, the execution time is significantly increased while working remotely by about half a second. User commentary indicates that the parquet flooring in the remote room was more slippery than the carpet in the co-located room which resulted in users being more careful in their movements. In addition to that, users had to drag the cable behind them in the remote room as it was lying on the floor. However, also the expert's indication times significantly increased in the remote setup even though experts' interaction remained the same. This indicates that users worked slower when connected remotely. However, the difference is small and users can cooperate independently of their physical location.

Furthermore, the results of the user study show that in general all gestures are suitable for the given task. Users' task performance is influenced by the gestures. However, the physical location of users does not impact gesture performance as indicated by the combined location*pointing technique results. All pointing gestures performed well enough for users to consider the ability to talk to each other as a surplus. No gesture seemed to have outperformed any of the other gestures in all aspects. The virtual hand gesture is easy to execute and reduces the indication time of the experts. However, the virtual pointer and target marker methods are overall significantly quicker. This might be because the user in the worker role could directly select the indicated cube and did not need to navigate around the expert. This result conforms with the conclusion of Bowman et al. [Bow01d] that all interaction techniques in VR have their strength and weaknesses and that there is no best technique.

The related work [Bor19, Góm16, Pod18a, Sou20, Syk20, Tan11, Wal09] confirms the similarity of a mediated co-located and remote collaboration. Qualitative ratings show that users feel equally co-present regardless of their actual, physical location. This supports H_B. However, even the co-located setup did not achieve full ratings of co-presence from every user which could be explained by the reduced environmental awareness in VR due to the immersion of the HMD and audio headset. Users surely focused and worked with the virtual avatar rather than the real human as she/he was not perceivable. The virtual avatar however is not realistic enough to transmit all bodily expressions that can be experienced through a real-world conversation.

4.4 Conclusion

Users can collaborate within VEs by being physically present in the same room or remotely via a network connection. A co-located collaboration benefits from a shared coordinate system to align the virtual worlds for all users. Currently, there exists no easy way to calibrate the different coordinate systems of a variety of devices. Therefore, an algorithm was developed and evaluated that calibrates two or more 6DOF devices. The algorithm has no special requirements for tracking sensors nor operating systems and supports low-power calibration. The devices need to be moved around the room on an interlinked trajectory through an easy-to-perform and low-effort slow rotation or spherical movement. The trajectory-based calibration offers higher precision when compared to camera-based calibration methods, i.e., SfM and marker tracking. The presented solution enables easy implementation of cross-platform co-located VR and AR applications.

When users collaborate in the same physical space, they need to be aware of the other parties to avoid collisions. Four different avatars were presented and evaluated in a user study for their capabilities and support in the context of collision avoidance. The results show that a basic representation of only the HMD and the controllers is not enough to achieve that goal. Further indications of the user's location, i.e., through an approximate cylindrical bounding shape or a full-body avatar, help to avoid collisions and boost users' feeling of safety. A shifted user representation, that adds a safety buffer, did not serve its purpose and leads to bad ratings and user insecurity.

When users start to explore the VE individually, their spatial relation changes. In this case, ghost-like avatars (see section 2.3.1) help to prevent collisions between users. In addition, digital voice communication via an audio headset is necessary to ensure that the direction of a user's voice matches the position of her/his virtual avatar. However, a VR HMD with spatial audio communication can also enable remote collaboration. The results of a user study show that co-located and latency-less remote collaboration differ only slightly regarding users' performance, preferences, and capability to collaborate when using a full-body avatar. A key factor for remote collaboration is the speed and latency of the network connection as it will probably influence the quality of the perceived co-presence. Further methods to compensate for network issues are discussed later in section 5.4. In addition to that, other techniques may be developed that close the gap between a co-located and remote collaboration even further, e.g., allow haptic contact between humans in remote setups.

When implementing multi-user collaboration within VEs, one should consider the following insights and design recommendations:

Visualize a full-body avatar: Visualizing users with a full-body avatar provides a natural collaboration, improves (co-)presence, and allows to avoid collisions. Approximate user movements can be calculated using inverse kinematics and few tracked objects. More advanced methods allow full-body tracking even without external sensors [Rho16]. However, even a low-fidelity full-body avatar can enable collaboration between users. The immersion into the virtual world and the feeling of co-presence might be increased even more with a more realistic avatar [Wei19a].

Immersive VR blocks out the human behind the virtual avatar: Based on the results of the user study and related work, collaboration within VEs seems to be independent of the physical location of users. As VR collaboration is still far away from replicating all aspects of real-world communication, it is questionable whether this is true. A photo-realistic avatar with accurate facial expressions and body movement, as well as physically accurate rendering of sounds and the environment are missing. It is unlikely that a remote VR collaboration as presented here will be able to replace a real-world face-to-face meeting, such as a contract negotiation, where every detail of the conversation counts. The similarities of co-located and remote collaboration might result from the fact that VR suppresses real-world impressions. Therefore, a VR user cannot perceive another human. Instead, users seem to engage with their virtual counterparts and focus solely on her/him. This might force users to extrapolate for missing information which may reduce performance and usability for both remote and co-located setups. Hence, collaborative applications might want to include other methods for users to express themselves to compensate for their avatar inaccuracies [Rot16], e.g., gestures, custom facial expressions/avatar appearances, or virtual drawings/emojis. Another approach might be to embrace the differences of the VE and to use them to increase task performance and usability.

5 Enhanced Virtual Worlds and Users

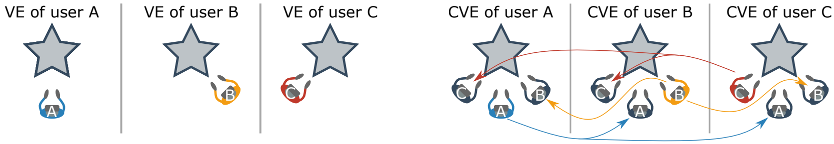
VR overwrites the user's senses to create immersive virtual worlds. These VEs can either replicate real-world experiences or let users perceive arbitrary spatial content. Without any effort, users can manipulate time and space. For example, objects can be manipulated regardless of their size and weight, or day and night can be switched at will. The other way around, the behavior of users can be manipulated by the virtual content. Redirected walking [Raz01] and touch [Koh09, Koh10] methods let users walk infinitely within a constrained physical space and experience haptic feedback using sparse objects. The redirection can also be applied to multi-user collaboration in CVEs. As the redirection is used in a social context, it is referenced to as *social redirection*. Social redirection allows altering the perception of the other users to e.g., alter social behavior [Rot18], provide users with a shared perspective [Sou19], or resolve technical limitations [Piu18]. One can imagine that content adjustments could also benefit the cooperation of users with different requirements, e.g., users with low vision. Each user may choose to adapt the virtual content (e.g., position, size, colors) for her/his needs. As the adaptations are only performed locally, several users would be able to engage in a transparent and inclusive cooperation.

First, this chapter presents an algorithm to implement social redirection supporting the manipulation of virtual objects and avatars. The algorithm is called personal perspective (PP) as it aims to provide individual users with an optimal perspective regarding their personal requirements. PP targets to improve task performance, usability, and collaboration. After that, the PP manipulation is applied to three use cases, i.e., rotation of a map for an upright view, a shift of user avatars to provide a shared perspective and face-to-face communication at the same time, and a replay feature to overcome short-term connection loss.

The content of this chapter is partially based on the following publications [Hop18d, Hop21].

5.1 Providing a Personal Perspective

Multi-user CVEs with VR HMD technology are composed of several separated worlds, one for each user, instead of one large world with all users in it. This is because generally, each HMD needs its own computer. To allow several users to work together in a shared environment, the different worlds need to be synchronized (see Fig. 5.1). Because of this, virtual worlds distinguish between two types of users: The *authority-user* is the user that works directly with the local machine and has authority over the local player object. The *connected-user* is a user that is replicated/synchronized over the network from her/his respective authority world and is displayed in the world of another authority-user. This means that each user is an authority-user on her/his local machine and at the same time represented as a connected-user on each of the machines of the other collaborators. Thus, the worlds are synchronized by sharing positional data of avatars and objects, the state of tools and menus, and other data such as voice between the different authority-users.



(a) Unsynchronized worlds.

(b) Synchronized worlds.

Figure 5.1: Three authority-users in their own VE (left) and their respective replicated connected-user avatars within a CVE (right) as viewed from each user's perspective.

The synchronization of the different virtual worlds poses an overhead when implementing a CVE. However, the separated worlds also allow modifying the experience for each user individually. The related work in section 2.3.3 states that on one hand, face-to-face interaction is preferred by users and improves task performance as it provides natural cues for intention and enables spatial partitioning. On the other hand, sharing a perspective with another user also increases task performance as it defines a referential context for the conversation and thereby provides a common understanding of the content. While both approaches have advantages, simultaneous usage is impossible in the real world. With the PP modifications presented below, it is possible to overcome the real-world limitations and manipulate the CVEs of multiple users individually based on their personal needs. For example, as commonly used by other systems presented in the related work, two or more users can work together on opposite sites of a table. This setup allows the collaborators to see the body expressions of the other party face-to-face while keeping an eye on the content at the same time. Moreover, this configuration is suitable for conversations with more than two people. The table allows the examination of different kinds of content like text, video, maps, or objects. Users can indicate certain locations on the table and talk about their findings. The problem with this setup in a real-life application is that only one person (or a small number of persons) is provided with the possibility to see the content upright. Other users around the table see the information rotated by maybe 90° or 180°. This makes it difficult to talk about relations within the content. Statements like ‘left of that’, ‘above’, ‘north of here’ have to be verbalized and interpreted in a group context to make sense. But, the translation of statements into a common context decreases task performance. A rotation of text labels might improve the situation but does not fully resolve this issue. Using PP, the CVE can be altered to turn the table content so that it is upright for each user. Thus, ‘left’ means ‘left’ for everybody.

5.1.1 Personal Transformations

To transform virtual objects into a PP, content should be able to translate, rotate and/or scale to give users a good perspective on it as well as to allow comfortable communicating with others. Independently translated content would allow two users to stand side by side but right in front of the same object. Different rotations allow users to see identical aligned content from different standpoints, e.g., when standing around a table. And scaling changes would allow them to work on differently sized objects, e.g., scaled down for better overview or scaled up for more visibility.

For the transformation of an object O into the personal view of a *user* A, an affine 4×4 matrix U_{OA} is applied. The personal transformation matrix U_{OA} allows altering the world of the user to her/his requirements by applying transformations such as translation, rotation, reflection, scale, or shear. To simplify and reduce the effects the matrix U can have, it can be decomposed as follows

$$U = \begin{pmatrix} & S \cdot R & & t \\ & & & \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{5.1}$$

R is a 3×3 rotation matrix that allows rotating content around the x, y, and z-axis (e.g., right-, up-, and forward-axis) about angles δ_x, δ_y, and δ_z:

$$\begin{aligned} R(\delta_x, \delta_y, \delta_z) &= R_z(\delta_z) \cdot R_y(\delta_y) \cdot R_x(\delta_x) \\ &= \begin{pmatrix} cos\ \delta_z & -sin\ \delta_z & 0 \\ sin\ \delta_z & cos\ \delta_z & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} cos\ \delta_y & 0 & sin\ \delta_y \\ 0 & 1 & 0 \\ -sin\ \delta_y & 0 & cos\ \delta_y \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & cos\ \delta_x & -sin\ \delta_x \\ 0 & sin\ \delta_x & cos\ \delta_x \end{pmatrix} \end{aligned} \tag{5.2}$$

S is a 3×3 diagonal matrix that allows to uniformly scale content by a factor of s:

$$S(s) = \begin{pmatrix} s & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & s \end{pmatrix} \tag{5.3}$$

Lastly, t is a 1×3 vector that allows to shift content by factors of t_x, t_y, and t_z along the x, y, and z-axis respectively:

$$t(t_x, t_y, t_z) = \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix} \tag{5.4}$$

For easier handling, t, R, and S matrices are trivially extendable to 4×4 matrices so that the following notation can be defined for an object O and a user A:

$$U_{OA} := t(t_x, t_y, t_z) \cdot R(\delta_x, \delta_y, \delta_z) \cdot S(s) \tag{5.5}$$

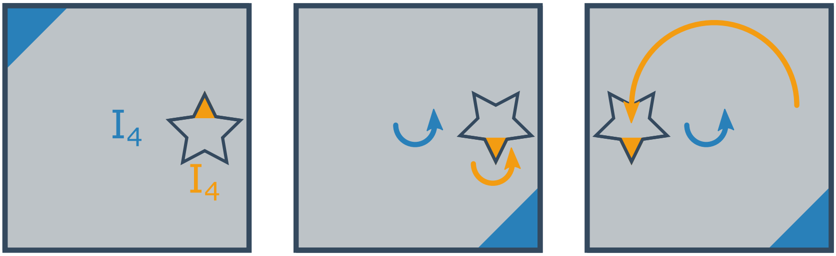
For example, U_{OA} equal to the identity matrix will result in an unaltered VE for *user A*. The matrix $U_{OA} = t(1{,}0{,}0) \cdot R(0{,}180^\circ{,}0) \cdot S(2)$ will double the size of the object, rotate it by 180° around the y-axis, and move it along the x-axis by a distance of 1. In a VE, the transformation of a virtual object O is described by a matrix M_O. M_O is the 4×4 model-matrix. It transforms the object O from its model-based local space into the world space of the VE. Using PP, the world-space transformation of a virtual object including the personal modifications can be calculated using the matrix P_{OA} with

$$P_{OA} := M_O \cdot U_{OA} \tag{5.6}$$

5.1.2 Content Modifiers

As described in the previous section, the personal transformation matrix U_{OA} can be applied to an object O to provide a *user A* with a personalized perspective. However, as virtual worlds contain not only one but many different virtual objects, it would be difficult and cumbersome to define different matrices U_{OA} for each of the objects. As objects are often semantically linked, it would be sufficient to define one personal transformation for a base object and let the other objects follow this transformation.

For example, a table T is rotated around its y-axis using a matrix $U_{TA} = t(0{,}0{,}0) \cdot R(0{,}180^\circ{,}0) \cdot S(1)$. An arbitrary object X that is placed on the table (see



(a) No modification: $U_{TA} = U_{XA} = I_4$

(b) Object-centric rotation: $U_{TA} = U_{XA} = R(0{,}180°, 0)$

(c) Table-centric rotation: $U_{TA} = R(0{,}180°, 0)$ $P_{XA} = M_T \cdot U_{TA} \cdot M_T^{-1} \cdot M_X$

Figure 5.2: Table T and object X (star) with different modifications applied.

Fig. 5.2a) needs to rotate around the table's center to keep its location relative to the table. To transform the object, it is not sufficient to use a matrix $U_{XA} = U_{TA}$ as it would result in an object that rotates around its own up-axis (see Fig. 5.2b). The transformation P_{XA} depends on the transformation of the table. To move the object accordingly, the following transformation can be used $P_{XA} = M_T \cdot U_{TA} \cdot M_T^{-1} \cdot M_X \cdot U_{XA}$. As the object does not perform any personal transformations, its matrix is the identity matrix $U_{XA} = I_4$. M_X transforms the object into world space. The combination of $M_T \cdot U_{TA} \cdot M_T^{-1}$ transforms the object from world into table space, applies the 180° rotation, and then transforms the object back into world space. As a result, the object correctly rotates around the table's center (see Fig. 5.2c).

As shown, the personal transformations of one object can become arbitrarily complex as interlinked movements of semantically dependent objects add up. To decrease the complexity of the object dependent personal transformations, the concept of a personal modifier is defined. Examples for a modifier are a table or a whiteboard. A personal modifier M, contains a personal modification U_{MA} for a *user* A. Every object that should follow the personal modification of M is transformed with respect to the personal transformation of the modifier. Analog to the object and table in the example above, any dependent object O

is transformed using the personal modifier M for a *user A* by

$$P_{OMA} := M_M \cdot U_{MA} \cdot {M_M}^{-1} \cdot M_O \tag{5.7}$$

The personal transformation matrix U_{MA} defines how virtual content should be adjusted to present an optimal view for a *user A*. Another *user B* might use a different personal transformation U_{MB} with $U_{MA} \neq U_{MB}$. To enable multi-user collaboration, a conversion between differently modified virtual spaces is needed. Besides transforming whole objects, it might be necessary to transform positions, rotations, or scale factors separately, e.g., to indicate a position with a virtual laser pointer. A location p_A in the world of a *user A* can be transformed to the VE of any other *user B* using the following two formulas:

$$x_l = {U_A}^{-1} \cdot {M_M}^{-1} \cdot x_A \quad = (M_M \ \cdot \ U_A)^{-1} \cdot x_A = {P_{MA}}^{-1} \cdot x_A \tag{5.8}$$

$x_A = (p_{A_x}, p_{A_y}, p_{A_z}, 1)^T$ is the 1×4 homogeneous coordinate of p_A. ${M_M}^{-1}$ and ${U_A}^{-1}$ transform the world position x_A into the local space of the object and revert the personal transformation that was applied through the modifier M for *user A*. As a result, x_l is the unmodified location in the local space of M. This coordinate can now be sent over the network connection to the other *user B*. In *B*'s world, the location now needs to be modified according to her/his personal perspective:

$$x_B = P_{MB} \cdot x_l \tag{5.9}$$

where x_l is the position that was sent over the network. P_{MB} modifies the location from local to world space and according to the personal modifier. $x_B = (x, y, z, w)^T$ is the resulting 1×4 modified position in the world space of *user B*. It can be converted to a 1×3 nonhomogeneous coordinate by calculating $p_B = (x_{B_x}/x_{B_w}, x_{B_y}/x_{B_w}, x_{B_z}/x_{B_w})^T$. Analog to a position, a direction vector d can be transformed with $x_A = (d_x, d_y, d_z, 0)^T$. It is transformed to a nonhomogeneous direction by discarding the w component.

To transform the rotation of an object, i.e., a quaternion q_A in world space of *user A*, the rotation $R(\delta_x, \delta_y, \delta_z)$ of the modifier M needs to be converted into a quaternion q_{MA}. The rotation can then be transformed into local space with $q_l = q_{MA}{}^{-1} * q_A$. In the VE of *user B*, the local rotation q_l then can be modified according to her/his personal perspective by the equation $q_B = q_{MB} * q_l$. Lastly, a rotation s_A that is modified with a scale factor s_{MA} (based on $S(s)$) can be transformed by calculating $s_l = s_A / s_{MA}$ and $s_B = s_l * s_{MB}$.

The calculations above transform objects from the personally modified world of a *user A* to the individually modified world of another *user B*. As virtual worlds can be large, multiple different personal modifiers might be located within the environment. To support multiple modifiers, each personal modifier defines an area of effect, e.g., defined by a bounding box or bounding sphere. Within this area of effect, every applicable object receives the respective personal modification. To transition between two or more (overlapping) areas, a smooth blend can be applied. This allows the use of different personal modifications and modification-free zones in one VE.

5.1.3 Modified User Avatars

The proposed modifications for the VE are not trivial and lead to additional problems. After applying different personal transformations on the virtual objects, the shared CVE has spatial inconsistencies between the different users. Therefore, valid position indications for one user deviate from the other users since the object's location, orientation, or size may have changed. Two approaches to resolve the inconsistencies are presented in the following sections. Before that, the underlying implementation to modify user avatars is discussed.

In current VR implementations, user avatars are often arranged as in Fig. 5.3a. A player object has a root or origin transformation with objects for the head, left hand, and right hand attached. The root object represents the origin of the tracking space that is defined by the VR system. The other objects follow the movement of the VR HMD and the two controllers. Additionally, further objects for a user avatar visualization or name tag can be attached to this hierarchy. To modify the user's representation and movements using PP, avatar modification objects are attached in-between the origin, head, left, and right hand objects and their respective child objects (see Fig. 5.3b). This allows to control the movement, rotation, and scale of the avatar by overriding the input of the user[1]. The implemented avatar modification objects inherit from the implementation of the modifier-adjusted objects as presented above. Because of this, the user's position, orientation, and scale with respect to the modifier are known within all CVEs.

The following two approaches for avatar modifications override the position of connected-user's avatars. But, based on the given information the avatars are adjusted to still point to the locations that the authority-user is currently highlighting.

(a) Unmodified avatar hierarchy.

(b) Modified avatar hierarchy.

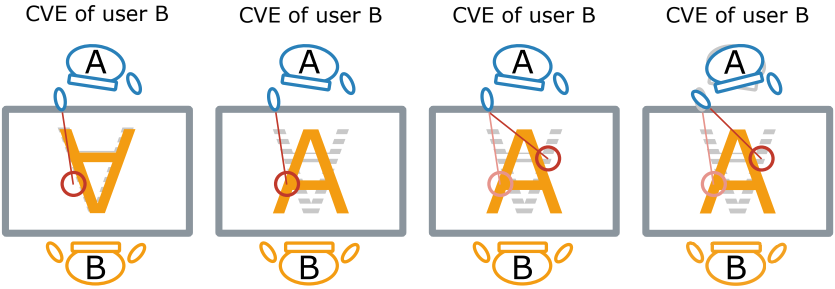
Figure 5.3: Additional transform objects are inserted into the avatar hierarchy to allow a modification independent of the user's movements.

5.2 Basic Avatar Modification

To validate the concept of a PP, a basic avatar modification was implemented. The avatar consists of the model of the HMD and the two controllers. Several systems within the related work use a ray-casting-based interaction method, as it allows to manipulate virtual objects, even from a distance, and provides

---

[1] Common game engines apply transformations to objects based on their hierarchy. Therefore, the four avatar modification objects are also detached from their respective parents to not receive any of the user's movement input.



(a) Unmodified/real-world table configuration.

(b) *User B*'s content is rotated by 180°. *User A*'s indicated position is now wrong.

(c) Correct location is highlighted but the avatar is not adjusted.

(d) Rotated content with modified avatar and indication. The correct location is marked.

Figure 5.4: The PP table as viewed by *user B* on the bottom side of the table. Note: The yellow content is only visible to the yellow *user B* and the gray-striped content is only visible to the top *user A*.

a tool to highlight positions. Therefore, a laser pointer tool was integrated for users to select, indicate, and manipulate virtual objects.

Two users are placed on opposite sites of a table (see Fig. 5.4a). To show upright content for both users, a PP modification is applied to the table which rotates the table's content towards each user (see Fig. 5.4b). However, as one user tries to indicate a location on the table using her/his laser pointer, the position of the indication relative to the content is not correct for the other user. To solve this, the point of intersection between the laser pointer and the table modifier is transformed using Eq. 5.8 and 5.9. Figure 5.4c shows the resulting indication with a correct spatial location. However, only adjusting the indication point does not suffice to fix the spatial manipulations. The avatars also need adjustment so that their attention and movements fit the altered indicated locations. To achieve this, a rotation is applied to the head and the controllers of each user (see Fig. 5.4d). A controller's forward orientation is adjusted to follow the direction of the altered laser pointer. The same is applied to the head of the user. The intersection between the forward view direction and the table is calculated. This intersection point is then transformed into the PP of the other user. Then, the head rotation is adjusted to focus on the transformed viewpoint. If the user is not looking or pointing at the table, no modifications are applied. A spherical linear interpolation is used to assure smooth transitions between the modified and unmodified avatar rotations.

5.2.1 Evaluation

Setup The table modifications were implemented using Unreal Engine 4 and two HTC Vive HMDs. As stated, the avatar of the users is basic and only shows the HMD and the two controllers. Figure 5.5 gives an outside view of the setup in the VE.

Figure 5.5: Outside view of the rotated PP table. Note: Laser pointer traced for print legibility.

Participants To test the proposed PP table modifications and to compare them to an unmodified environment, a within-subject user study with 16 participants was conducted. The participants consisted of 13 male and 3 female subjects with an average age of M=32.4 years (SD=11.0). On a scale from 0 (none) to 4 (very high), participants rated their experience with VR and with maps as M=1.500 (SD=1.211) and M=2.813 (SD=1.167) respectively.

Tasks and Procedures Subjects shared the VE with a test supervisor that stood on the other side of the table. Before the test, users learned the interactions and accustomed themselves with a training task. The subjects were asked different questions about a map that was displayed on the table's surface. The test supervisor read the questions and pointed to different locations with a laser pointer. The questions were basic and asked users to describe the relation of locations, the directions of a path, and to find a shortest path. Users answered five training and eight timed questions for each condition. Most of the questions required users to answer in cardinal directions. Because of that, two compasses are positioned on the table, one on each side. Participants performed the tasks with two conditions. Once, with a real-world replication where the test supervisor had a north-oriented map and the subject saw the map upside down. The other time, the map was north-oriented for both users and the PP modifications were applied. Figure 5.6 shows the in-game views of both users with the two map orientations. To eliminate training or fatigue effects in the evaluation of the collective results, the order of the real-world and PP world was mixed. Furthermore, two different maps were used (tourist maps of London and Tokyo). The two map rotations and two maps result in four different combinations.

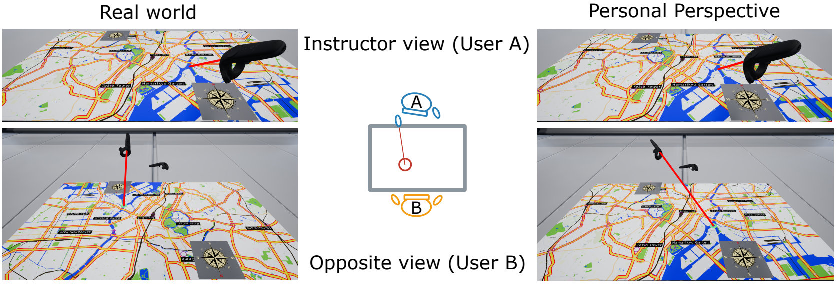


Figure 5.6: Comparison of real-world and PP viewpoints. In the real-world case, the table content is upside down for the user on the opposite side of the table. With applied modifications, both users can see the content upright in the PP case. Note: Laser pointer traced for print legibility.

Measurements The upright content representation should make it easier for users to accomplish the presented tasks. Therefore, the following hypotheses were formulated for the results of the evaluation:

H_A Task performance is increased with the PP modification compared to the unmodified table representation.

H_B The workload of users is decreased while using the PP table.

During the user study, quantitative and qualitative data were collected. The performance of the users was quantified with time and error measurements. After each of the two tasks, users were prompted with a questionnaire in VR. The questionnaire contained six questions from the RTLX. Additionally, users were asked how natural the conversation with the test supervisor felt. Furthermore, users were asked how much they depended on the compass while orienting themselves on the map. Users rated the different questions on a five-point Likert scale ranging from 0 (low) to 4 (high).

5.2.2 Results

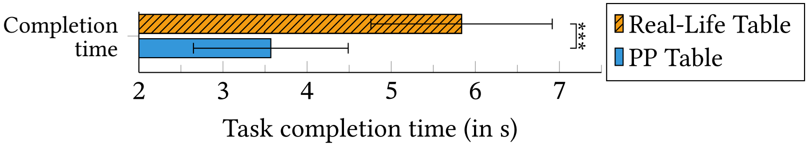


Figure 5.7: Task completion time. Results as average value with 95% confidence intervals and asterisks indicating significant differences.

Quantitative Results Figure 5.7 shows the quantitative results of the user study. Participants are about 1.6 times slower with the upside-down map M=5.837 (SD=6.221) compared to the north-oriented map M=3.569 (SD=5.310). The differing question complexity leads to high standard deviations for the real-world and PP table respectively. A paired t-test shows a significant difference between the conditions (t=5.394, p<.001, N=128) with a medium effect size (Pearson's r=.432).

Overall, only five errors were made with the upside-down table and one error with the modified PP table. During the south-up map tests, however, users often switched answers (“west… no, I mean east”), which in return also explains the higher response times.

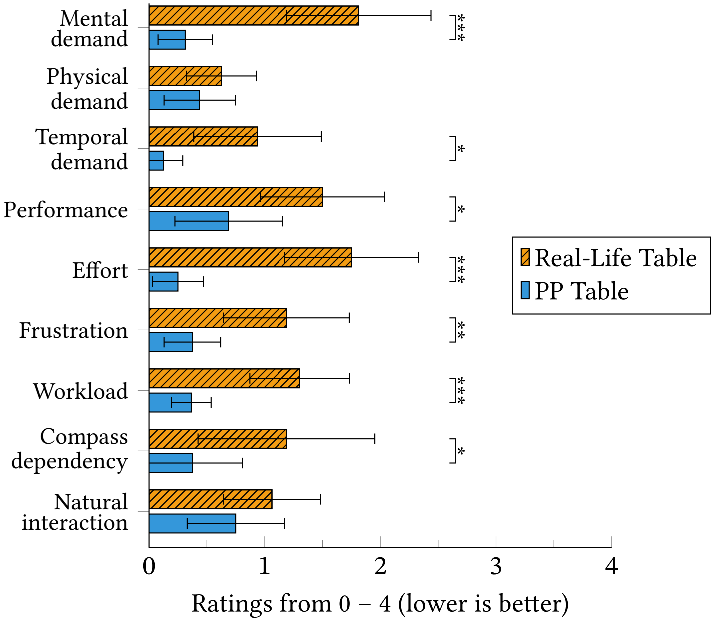


Figure 5.8: User ratings. Results as average value with 95% confidence intervals and asterisks indicating significant differences.

Qualitative Results Figure 5.8 displays the ratings of the custom questionnaires. The mental demand with the modified, i.e., upright, PP map is M=0.313 (SD=0.479), whereas it is M=1.813 (SD=1.276) with the south-up map rotation. A WSR test shows that the difference is significant (z=-3.235, p=.001, N=16) with a large effect size (Pearson’s r=.809). The real-life and modified tables’ physical demand is M=0.625 (SD=0.619) and M=0.438 (SD=0.629) respectively. According to a WSR test, the difference in physical demand

was rated with no significant difference (z=-1.342, p=.180, N=16). The measured temporal demand might be more of a psychological effect since users did not have a time limit. Still, the PP table M=0.125 (SD=0.342) induced significantly less temporal demand than the unmodified table M=0.938 (SD=1.124) based on a WSR test (z=-2.392, p=.017, N=16). The effect has a large size (Pearson's r=.598). Users rated their own performance in the user study as high with M=1.500 (SD=1.095) and M=0.688 (SD=0.946) for the real-world and PP table respectively. Their assessment of performance is significantly higher in the upright rotated case (z=-2.289, p=.022, N=16) with a large effect size (Pearson's r=.572). The effort of the participants to reach their performance result is M=1.750 (SD=1.183) with the normal map and M=0.250 (SD=0.447) with the modified map orientation. A WSR test shows a significant difference (z=-3.256, p=.001, N=16) with a large effect size (Pearson's r=.814). Participants experienced low frustration for the real-life map M=1.188 (SD=1.109) and even lower for the PP map M=0.375 (SD=0.500). A WSR test shows the difference is significant (z=-2.565, p=.010, N=16) with a large effect size (Pearson's r=.641). The overall workload is M=1.302 (SD=0.878) for the real-life map and M=0.365 (SD=0.351) for the PP map. According to a WSR test, the difference is significant (z=-3.327, p=.001, N=16) with a large effect size (Pearson's r=.832).
The dependency on the compass while answering the questions fluctuated between users. On a scale from 0 (low) to 4 (high), it is higher with the real-world table M=1.188 (SD=1.559) than with the PP table M=0.375 (SD=0.885) because of the unfamiliar map rotation. A WSR test shows the difference is significant (z=-2.392, p=.017, N=16) with a large effect size (Pearson's r=.598). Users rated their interaction with the test supervisor on a scale from 0 (very natural) to 4 (not natural). The interaction feels natural with M=1.063 (SD=0.854) for the real-life table and M=0.750 (SD=0.856) for the PP table. Based on a WSR test, the difference is not significant (z=-1.186, p=.236, N=16).

After both table conditions, users assessed if they could solve the map tasks better with the north-oriented map than with the upside-down map as M=0.812 (SD=1.223) on a scale from 0 (yes, totally) to 4 (no, not at all). The final question asked if participants noticed that the worlds that they and the test supervisor saw were sometimes not identical. Only one subject noticed that the movements of the avatar with the PP rotated table sometimes did not fit the movement of the indicated position. Concluding comments of four participants highlight that the texts in the VE are difficult to read. The main reason for this is the low resolution of the HMD. For now, bigger fonts or smaller distances between the user and the text can be used to compensate for this.

5.2.3 Discussion

As the user study shows, it is beneficial for users to work with a familiar representation of content. This supports H_A. Working with upside-down content takes longer with an average time increase of about 64%. It is conceivable that this extra time contains mental rotations and error corrections for an initially wrong, immediate response that is based on a familiar view. As presented, participants rated that the north-oriented map is much better to work with than the upside-down map. This can be interpreted as a clear advantage for the familiar north orientation. The results are consistent with the related work [Ann91, Byr02, Fer92, Gra84]. Moreover, the need to gather a mutual understanding of the virtual world by building a common mental representation is not required. Through PP, users can share a common view of the content and thereby increase their task performance. The participants of the user study have a high experience with maps and it might not be too unfamiliar for them to view a map upside down. Still, there is a significant advantage for the north-oriented map. Less experienced users might have an even bigger benefit. The PP table findings show that users are hardly dependent on the displayed compass with north facing upward. The modification of PP results in less demand, effort, and frustration and decreases the workload significantly which supports H_B. The cooperation also feels natural to users.

Only one user noticed, that the two VEs of the expert and worker were not identical. Roth et al. [Rot18] also found the number of modification detections is low. This user hinted that the movements of the other avatar were inconsistent with respect to the movement of the indicated positions. Figure 5.9 explains the mismatch between avatar movement and indication. The movement-indication mismatch originates from purely rotational modifications and lack of translational adjustments of the avatar.

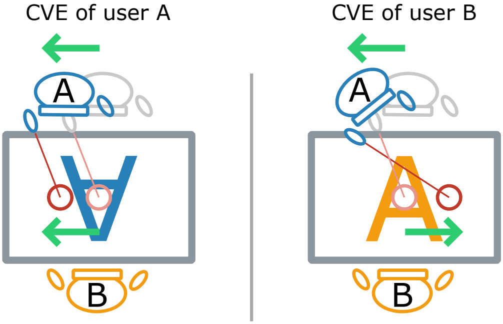


Figure 5.9: The movement-indication mismatch. On the left (view from supervisor, *user A*), the movement of the test supervisor and the indication are consistent. On the right (view from subject, *user B*), the avatar of the supervisor and the indicated position move in opposite directions.

Only adjusting the rotation of avatar objects leads to further difficulties besides the movement-indication mismatch. Indications in a rotated, translated, and scaled world can be transformed rather simply into the PP of other users. In contrast, direct contact with virtual objects creates some difficulties. If *user A* touches an object in her/his world, the transformed object in the world of *user B* might have a different position. It is clear that a movement of that object by *A* results in a movement in the world of *B*. The problem is that *A* is moving but not touching the object in the world seen by *B*. The reason for that is the adjustment of *A*'s avatar, which only uses rotation. This inconsistency might be solved by introducing a ghost hand, drawing a connection between the object and *user A*, or extending *A*'s reach with 'inspector gadget hands' as in [Pou96] so that there is a clear indication of who is moving the object.

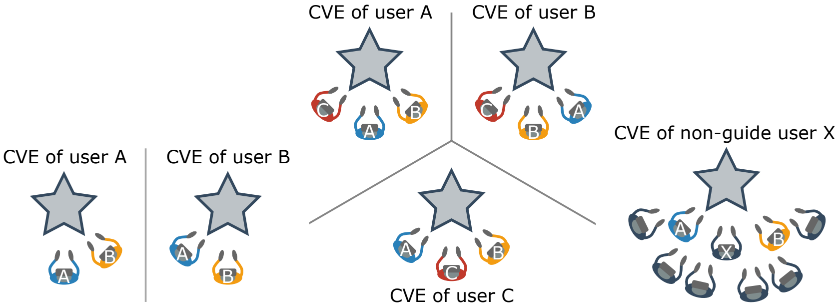
The movement-indication mismatch and the missing direct interaction may disturb the VR experience of users. But, with the presented task, the mismatch is often overlooked and does not draw attention to itself. Content rotation is a valid improvement that increases work productivity and decreases error probability.

5.3 Extended Avatar Modification

The evaluation of the basic avatar modification from the previous section shows that a rotational modification of users' avatars improves task performance significantly but is lacking functionalities such as direct object manipulation, i.e., grabbing. Therefore, this section introduces an extension of the previous avatar adjustments to a full-body avatar to resolve the presented issues.
Wong and Gutwin [Won10] show that deictic pointing in CVEs is accurate and only slightly worse than real-world pointing. However, later they found that a laser pointer can increase the pointing performance and is preferred by users [Won14]. Therefore, users can switch between a direct (grab) and a remote interaction (laser pointer) [Haa06] as needed. The length of the laser pointer is adjustable to allow a fishing reel style manipulation of objects [Bow97a].

The presented method provides users with a shifted but shared perspective (ShiSha). To provide all users with the same perspective, each authority-user stands at the ideal location, e.g., in front of an object of interest, in her/his authority world. To avoid overlapping with the avatars of connected-users, their avatars are shifted to different locations, e.g., to the left or right of the authority-user. This allows a shared perspective interaction with the advantages of side-by-side/face-to-face communication.
The applied shift changes the pointing and grabbing targets of connected-users as the actual position of an authority-user and the shifted position of her/his connected-user's avatar do not match. To overcome the resulting inconsistencies, a distance-based redirected touch interaction approach is used, which adjusts head and hand movements.

Two Users or More The ShiSha technique targets remote support or training scenarios which often involve two users, e.g., an expert and a worker. In these contexts, it is beneficial for the expert to have the same perspective as the worker to be able to help based on a shared understanding of the virtual space. To support this, all users are positioned at the same location. Without ShiSha, the bodies of users would overlap and users might not be able to perceive the avatars of the other collaborators. But, as the related work shows, it is beneficial if users see each other to communicate via gestures or other expressions. Therefore, ShiSha shifts connected-users' avatars to a different position to enable face-to-face interaction in a side-by-side configuration.



(a) Two-user ShiSha setup **(b)** Three-user ShiSha setup **(c)** Nine-user ShiSha setup. Note: only one of nine CVEs is displayed.

Figure 5.10: The authority-user stands in front of the object of interest (star). The connected-users are shifted to the side by a rotation around the object to avoid overlapping avatars.

For example, in the case of two users, each user has authority over one virtual world. ShiSha applies a shift to move the connected-user's avatars to the side. As Fig. 5.10a shows, in the CVE of *user A*, *user B* is shifted to the right, and in the CVE of *user B*, *user A* is shifted to the left. Yet, complex content or larger classrooms may demand collaboration with three users or more. With three users, there exist three virtual worlds (see Fig. 5.10b). Again, each user is positioned right in front of the object of interest. The connected-users are shifted to the left and right. Larger groups might follow an asymmetric role allocation with one or a few persons taking the role of a guide/presenter/teacher. Therefore, all guides are shifted right next to the authority-user so that she/he can follow the guide's explanations (see Fig. 5.10c). All other users are positioned further to the side or behind her/him. The locations of the users are of course variable and can be changed dynamically, i.e., two users switching position if necessary. This allows for dynamic role allocations and consideration of users' priorities, e.g., higher priority if somebody asks a question.

In the case of a two-user interaction with ShiSha, the shift preserves the relative locations of both users and does not need any additional modifications, apart from the avatar redirection explained below. However, three or more users and dynamic position changes induce a mismatch of the relative locations between the users. Figure 5.10b illustrates this. In the CVE of *user B*, *user A* is positioned to the right of *user B*. But, in the CVE of *A*, *B* is also located to the right of *A*. This means that in the world of both users, the respective other user is shifted to the right. Turning right to face this user would lead to turning away from the user in her/his world. To overcome this, the ShiSha modification can be extended with a look-at rotation that compensates for between-user inconsistencies. This is done by detecting if *user A* is facing the other *user B*. If so, the connected-avatar of *A* in the world of *user B* is rotated around the up-axis to face the looked-at person, in this case, *user B*. If the user is not looking at any user, no look-at rotation is applied. This assures that a connected-user's avatar always faces the same person that the respective authority-user is looking at and allows more than two users to benefit from the ShiSha adjustment.

Avatar Modification ShiSha applies no shift or redirection to the authority-user. Only the avatars of connected-users are shifted to the side and positioned beside the authority-user around the object of interest, e.g., a table or a machine. This is implemented by rotating the users around the object's center by an angle α (see Fig. 5.11).

Due to the change of position and rotation, grabbing and pointing locations of the connected-users are also changed. ShiSha transforms the shifted avatars

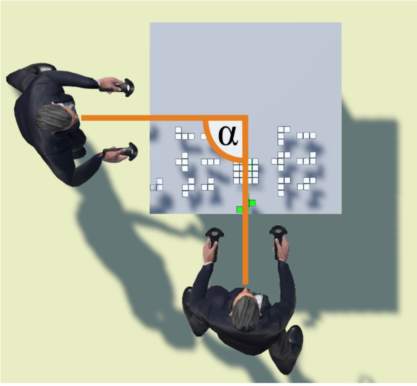


Figure 5.11: The connected-user's avatar is shifted around the object of interest by an angle α.

using redirected touch interaction to compensate for the resulting inconsistencies. Linear interpolation and spherical linear interpolation between the unmodified/real and the shifted location modifies the position and the rotation of the connected-users' heads and hands. The interpolation is distance-based and assures that if a user is close to an object, the shifted avatar will be close to the object also.

During the fishing reel style laser pointer interaction, the shifted and redirected remote avatar of *user A* in the world of *user B* also uses a laser pointer. However, when using direct interaction, i.e., touching or grabbing, objects that are within reach for *user A*'s touch in the world of *A* can become unreachable for *A*'s avatar in the world of *user B*. Thus, the arms of *A*'s modified avatar could stretch in an unnatural way to reach the object. To maintain bodily consistency and not decrease co-presence due to unrealistic arm-stretching [Bai05, Zib18], the extension of the arms was limited to 75 cm, which is slightly less than the average adult arm length [Živ03]. A connection is needed to bridge the gap between the limited hand location and the interpolated target location. This could be done using some kind of beam or a mechanism like 'the force'. The connection was visualized similarly to the provided fishing reel manipulation technique. If the arm length reaches the threshold, the laser pointer on the respective arm is automatically activated and points to the interpolated location. The rotation of the hand is adjusted smoothly to follow the direction of the laser pointer. Yet, a pointing hand conveys more information due to its orientation than the tip of a laser pointer. To convey the pointed-at direction, the laser was curved using a bézier curve (see Fig. 5.12). The curved pointer, therefore, visualizes the direct grab position and rotation of a connected-user for an out-of-reach object.

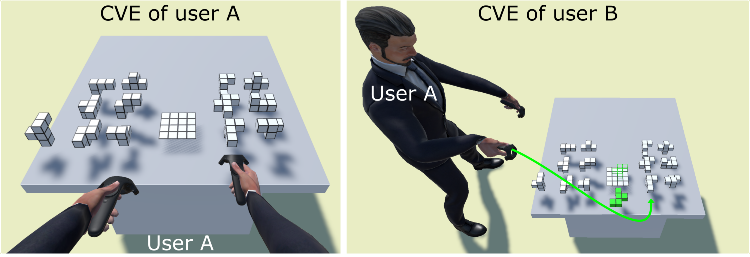


Figure 5.12: *User A* reaches for an object (left). The redirected avatar of *A* cannot reach the object without unnatural arm-stretching (right). Therefore, ShiSha limits the length of the arm and activates the curved laser pointer for *A*'s avatar. Note: Laser pointer traced for print legibility.

With ShiSha, two or more users can share the ideal perspective. But, this ideal point of view might not always be obvious to all users. One user might want to direct the perspective of the others and highlight a certain part of an object. In this case, one user can guide others by enforcing a certain point of view on them (position-guiding). Only in this instance, the authority-user is modified in that she/he is teleported and oriented to imitate the perspective of the guide.

5.3.1 Evaluation

Setup To evaluate the extended method and to show a different use case for the PP technique, the table setup was slightly altered for ShiSha. Instead of rotating the object of interest towards the users, the users themselves are shifted. Conceptually this results in the same changes as before. However, as the users and not the virtual content is moved, the surroundings of the VE keep their relative positions with respect to the personal modifier.

To evaluate the basic concept of ShiSha, a user study was conducted with dyads performing a collaborative task. Besides ShiSha, the method of Vishnu [Le 16a] (see Fig.5.13 and section 2.3.2) was implemented since it is one of the few techniques that allows two users to share the same view with high levels of performance and a good feeling of co-presence. With Vishnu, two users stand at the same location and see one additional pair of arms from their partner next to their own arms. The two techniques used in the user study are comparable to the same-side and around-the-table configurations of Tang et al. [Tan10] with the difference that with ShiSha users share the same perspective even though standing side-by-side.

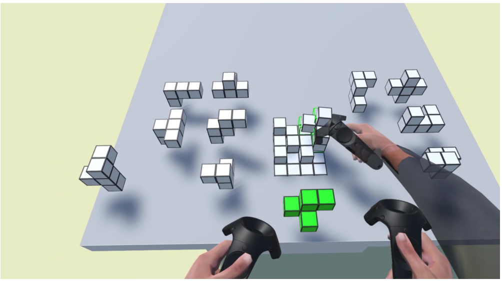


Figure 5.13: The baseline method Vishnu.

The user study was performed with dyads of VR systems. As opposed to a VR-AR or AR-AR collaboration, VR-VR allows full control over the CVE and removes distractions from the surroundings. Also, the VR-VR setup allows to compare ShiSha and Vishnu with two users inside the same room which is comparable to working remotely (see section 4.3). The position-guiding feature was used to provide both users with the same perspective. Based on the findings of Sommer [Som59], a shift angle of α=90° was used to position the users next to each other at the configuration depicted in Fig. 5.11. Since the users were standing next to each other in the same room, they could talk directly to each other. Users were permitted to move around to adjust their perspective to better understand the object's shape. However, they were asked not to leave the front side of the table, as it represents the ideal location to execute the task.

The user study was conducted using two HTC Vive Pro VR HMDs. The two computers were connected using a 1 Gbps Ethernet connection. The CVE was developed using Unity. Users were represented with a full-body avatar with a semi-realistic appearance. However, no facial expressions nor a female version were available.

Participants Through a poll, 32 volunteers (22 male, 10 female, M=32.6 years, SD=10.4) were recruited. Rated on a seven-point Likert scale from 1 (novice) to 7 (expert), users self-assessed their experience with VR as M=3.688 (SD=1.693). Only three users had never used a VR system before, 22 use VR at least once a year, three at least once a month, two at least once a weak, and two daily. Their self-rated ability for spatial thinking was measured on a scale from 1 (low) to 7 (high) and yielded M=5.156 (SD=1.167).

Tasks and Procedures ShiSha is not limited to a turn-based task but is also applicable to cooperative object manipulation. However, to set up the dyads of the user study in expert and worker roles, an assembly task was chosen that contained additional knowledge for only one user which needs to be transferred to the other user. Participants unknowingly chose their role based on their positioning while filling out the initial questionnaire (left desk ↦ worker, right desk ↦ expert). The roles were maintained throughout the whole experiment. Assembly tasks are commonly used in the evaluation of collaborative systems in either symmetrical, e.g., [Rob03, Ste03, Wid00], or asymmetrical roles, e.g., [Ant18, Fus03a, Ger04, Hua13a, Kir06, Kol18, Sal09, Teo18]. Just as in previous work, a generic block assembly task was selected that employs similar actions as more complex assemblies [Kir06]. The Bedlam cube [Chl05] is a 4×4×4 cube that consists of 13 distinct pieces. The cube can be solved in 19,186 different ways and serves as an ideal puzzle for the evaluation since the worker would not be able to predict the solution but rely on the expert's information. The 13 pieces of the puzzle were placed on the table around a

4×4 platform, i.e., the base of the finished puzzle. To provide the expert user with prior knowledge, she/he is presented with a step-by-step construction plan of the Bedlam cube. A green copy of a puzzle piece indicates the next puzzle piece that needs to be used and a transparent green highlight shows its target position (see Fig. 5.14). The user in the expert role was asked to identify the current piece and then advise the worker to place that piece in the targeted position. Upon release, puzzle pieces snap to a grid to allow faster positioning to remove any problems regarding precise placing. To prevent the expert from solving the puzzle on her/his own, the expert is only able to rotate, but not to move the puzzle pieces.

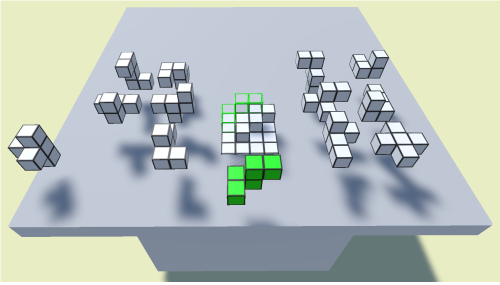


Figure 5.14: Puzzle as seen by the expert, with a copy of the target piece (solid green) and the target location (green outline).

In the user study, the participants were asked to solve a total of seven Bedlam cube puzzles. During the first puzzle, users could familiarize themselves with the controls of the system, i.e., grabbing objects with direct and remote control. Afterward, dyads performed three puzzles for each of the two techniques ShiSha and Vishnu. To compensate for any effects of learning and fatigue, the two techniques were performed in mixed order. Since the number of groups in the user study is a multiple of two, the results are counterbalanced. In between the puzzles, dyads had up to two minutes of debrief time. During this time, they could optimize their strategy for the next puzzle, e.g., update their vocabulary for the positioning or the pieces. The debrief was aborted if dyads had nothing more to discuss and wanted to continue. Speech communication was free from restrictions during the whole time of the experiment. The experiment took about one hour, including introduction and questionnaires.

At the beginning of the user study, all participants filled out a demographic questionnaire. After the three puzzles of technique one, participants took off their VR HMD and filled out questionnaires regarding performance, usability, and social presence. Similarly, users took of their VR HMD and filled out the same questionnaires after technique two's three puzzles. A final questionnaire asked the participants to compare the two techniques.

Measurements ShiSha allows users to engage in a face-to-face conversation to facilitate a higher co-presence and better teamwork. However, engaging in a face-to-face conversation may reduce the focus on the presented task when using ShiSha. Also, the redirection might induce some issues regarding spatial referencing and cause communication errors. At the same time, a shared perspective [Fus00, Ger04] and visual cues from other avatars [Fus03b, Gar01, Piu18] may improve performance. Overall, ShiSha aims towards balancing usability, task performance, feeling of teamwork, and perceived co-presence. Therefore, the following hypotheses were formulated for the results of the evaluation:

H_A ShiSha and Vishnu have high usability as both methods aim to improve usability during collaboration.

H_B ShiSha has an equal or higher task performance than Vishnu. As Vishnu is designed to increase task performance, ShiSha targets to have at least the same effectiveness.

H_C Social presence, especially co-presence, is higher with ShiSha than with Vishnu because ShiSha allows users to see each other.

The two techniques ShiSha and Vishnu were compared using objective and subjective measurements. Objective measurements included the time taken to solve the puzzles and the time spend looking at the other user. Looking at the other user is an important part of the conversation. Since eye-tracking is not available with the given VR HMDs, head gaze was used. Head orientation has a large contribution to eye gaze direction and attention focus [Sti02] and gives an indication, whether the participants engage in face-to-face interaction. It was defined that a user is looking at the other user if the angle between her/his forward head direction and the direction to the head of the other user is lower than 18°. At roughly 2 meters distance between the users, this yields a look-at width of 0.65 cm, which is slightly larger than the average shoulder width [Sto70] to accommodate for eye movement. This value is therefore not exact. For the subjective measures, participants were asked to fill out three questionnaires, i.e., RTLX, UEQ-S, and NMSPM. At last, users were presented with a question regarding their preference for the technique on a seven-point Likert scale. Also, participants were asked to provide written feedback regarding their experience with the techniques.

5.3.2 Results

Quantitative Results On average, participants took M=283.849 seconds (SD=108.525) to complete one puzzle with ShiSha and M=290.959 seconds (SD=150.381) with Vishnu. Given the continuous completion time and dependent measurements, a two-way repeated-measures ANOVA with the technique (i.e., Vishnu and ShiSha) and repetition (i.e., 1, 2, 3) as factors was used (see Fig. 5.15). Mauchly's test indicates no violation of the assumption of sphericity for repetition ($\chi^2(2)=0.218$, p=.897) nor technique*repetition ($\chi^2(2)=0.471$, p=.790) (not calculated for the technique as sphericity is always met for two levels). There are no significant differences for technique (F(1,15)=0.086, p=.773, N=16), repetition (F(2,30)=1.859, p=.173, N=16), nor technique*repetition (F(2,30)=0.253, p=.778, N=16).

The look-at times below are measured with the ShiSha technique. With Vishnu, it is not possible to face the other user. Experts looked M=106.383 seconds (SD=69.534, 8.6% of the total duration of the experiment) at the user in the worker role. The workers looked M=190.976 seconds (SD=97.491, 14.8% of the total duration of the experiment) at the user in the expert role.

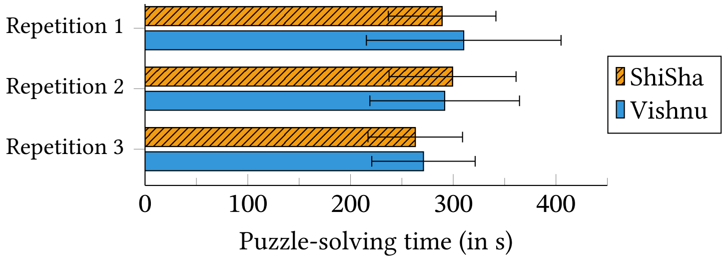


Figure 5.15: Task completion time per puzzle. Results as average value with 95% confidence intervals.

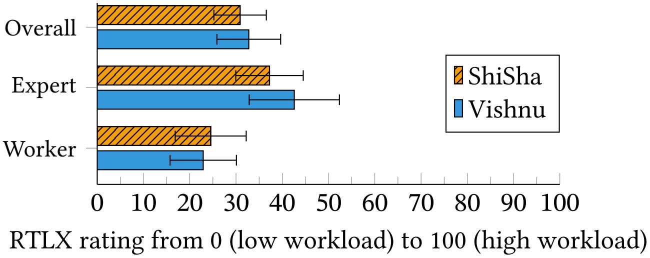


Figure 5.16: RTLX scores overall and grouped by expert and worker roles. Results as average value with 95% confidence intervals.

Qualitative Results Overall, ShiSha achieves an RTLX workload score of M=30.885 (SD=16.356) and Vishnu of M=32.760 (SD=19.879) (see Fig. 5.16). According to WSR, there is no significant difference between the two scores (z=-0.617, p=.537, N=32). The user's workload is slightly higher for the expert with M=37.240 (SD=14.884) for ShiSha and M=42.604 (SD=19.898) for Vishnu than for the worker with M=24.531 (SD=15.658) for ShiSha and M=22.917 (SD=14.631) for Vishnu. Again, the differences are not significant for the expert (z=-1.421, p=.163, N=16) nor the worker (z=-1.005, p=.315, N=16).

The results of the UEQ-S are presented in Fig. 5.17. Overall, ShiSha achieves a UEQ-S score of M=1.430 (SD=0.769) and Vishnu of M=1.434 (SD=0.774). There is no significant difference between the two scores (z=-0.103, p=.918, N=32). For the pragmatic quality of the techniques, ShiSha achieves a score

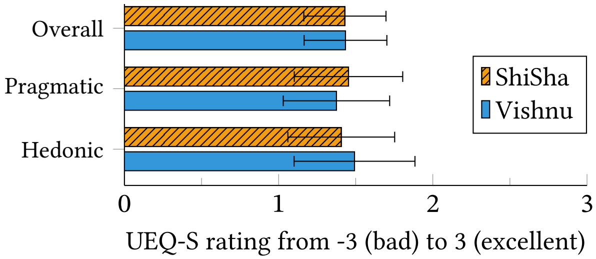


Figure 5.17: UEQ-S scores. Results as average value with 95% confidence intervals.

of M=1.453 (SD=1.015) and Vishnu of M=1.375 (SD=0.996). The difference is not significant (z=-0.354, p=.723, N=32). For the hedonic quality of the techniques, ShiSha achieves a score of M=1.406 (SD=0.999) and Vishnu of M=1.492 (SD=1.131). Again, there is no significant difference between the two scores (z=-0.392, p=.695, N=32).

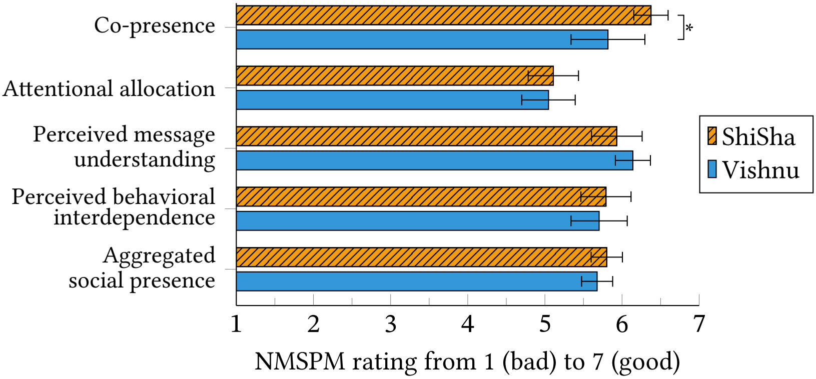


Figure 5.18: NMSPM scores. Results as average value with 95% confidence intervals and asterisks indicating significant differences.

The results of the NMSPM are presented in Fig. 5.18. For the feeling of co-presence, users rated their experiences with the techniques ShiSha and Vishnu as M=6.375 (SD=0.641) and M=5.818 (SD=1.381) respectively. The difference between the two is significant (z=-2.275, p=.023, N=32) with a medium effect size (Pearson's r=.402). The attentional allocation was ranked as M=5.109 (SD=0.941) for ShiSha and M=5.047 (SD=0.999) for Vishnu. There is no significant difference between the two scores (z=-0.444, p=.657, N=32). Users perceived message understanding yielded M=5.932 (SD=0.947) for ShiSha and M=6.141 (SD=0.656) for Vishnu. The difference is not significant (z=-1.021, p=.307, N=32). The perceived behavioral interdependence was ranked by the participants with M=5.792 (SD=0.938) for ShiSha and M=5.703 (SD=1.051) for Vishnu. The ranking is not significantly different (z=-0.337, p=.736, N=32). Overall, the aggregated social presence of the technique ShiSha is M=5.802 (SD=0.585) and M=5.677 (SD=0.581) for Vishnu. There is no significant difference between the overall scores (z=-1.585, p=.113, N=32).

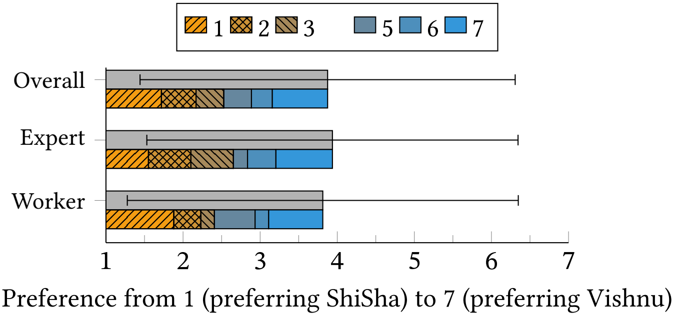


Figure 5.19: User preferences for the techniques. Results as average value with SD error bars and share in user votes as a color gradient. Note: No participant chose value 4 (no preference).

Users do not have a clear preference for either ShiSha or Vishnu, as Fig. 5.19 shows. On a scale from 1 (preferring ShiSha) to 7 (preferring Vishnu), the overall rating is M=3.875 (SD=2.433). Users in expert roles and users in worker roles had a similar preference with M=3.938 (SD=2.407) and M=3.813 (SD=2.536) respectively. As the share of user votes in Fig. 5.19 illustrates, the preference was evenly distributed across the possible answers, except for a neutral preference (4), which was never chosen.

During the user study, users were really engaged in the task and steadily tried to improve their collaboration. Because of the training and the simple input, even inexperienced VR users did not have any problems with the interaction. Each group developed a task-oriented language at the start and kept it even after the technique had changed. In the first few puzzles, users worked on efficient communication to indicate puzzle pieces and their target position. Puzzle pieces were mostly specified using a pointing hand gesture or the laser pointer. The target position was often described as left/right, up/down, or front/back. However, with increasing progress, workers could sometimes see the target position of the piece and no communication was necessary. In later puzzles, workers often tried to help the expert to search for the next puzzle piece since they had to wait on her/him. The expert described the puzzle pieces using characteristic shape properties, such as flat or 3D, the number of blocks, figurative meaning (e.g., “cross”, “stairs”), or “has 3 blocks in a row”.

15 users indicated in the written feedback that the collaboration with ShiSha is “natural” and “analog to interacting in the real world” and can be “experienced as teamwork”. “Here you had the feeling that communication worked better and that you worked together”. Moreover, 20 participants wrote that the face-to-face interaction of ShiSha allows them to “follow and observe the other user and [her/his] movements”. Three users wrote that they had problems with the bend laser pointer because the “end of the laser pointer was floating in the air”, that it was “difficult to see, what [she/he] was pointing at”, and that it “distracted [their] attention from the task”. Also, “sometimes the arms [had] an unnatural posture” and the avatar’s “movements were not exactly visualized as intended”. Two users noted that the avatar is missing facial expressions for a “realistic conversation” which is “noticeable during the discussion”. One found that “the avatar was annoying”. However, “the abstraction of the avatar’s appearance is ok for solving the task”. One user notes that she/he “could not remember details of movement, because [she/he] was concentrating on the task”. “The avatar does not look like my partner but that does not influence the interaction”. Two users note that it was “not disturbing, but a little irritating that [although the other user was standing beside me, she/he] had the same perspective”. One user was bothered that the other

user “put blocks in [her/his] way or too far away”. One user wrote: “I liked that I could map the voice of my partner to a person in VR”.

With Vishnu, five users wrote that they could follow the arm movements of the other’s avatar. Seven users wrote that the arms were blocking their view and eight users wrote that the “extra pair of arms” is irritating. Users wrote that the “arm movement was unnatural”, that “flailing hands came into [their] field of vision” and that they “disliked the fact that [they] had the arms of [their] partner in front of [them]”. Six users described that the other avatar was not noticeable as the “partner’s actions seemed to come out of nowhere” and they “hardly noticed the way [the partner] moved”. “I did not have the feeling that another person was carrying out the action”. Three users described their experience with Vishnu as lonely. “I felt like I was doing everything alone”. Three users had the feeling that the other user was standing behind them and two expressed that they felt watched. “[Vishnu] was psychologically uncomfortable, because you felt that your partner was getting too close”.

Seven users did not notice large differences regarding the interaction between Vishnu and ShiSha, as they “found no noteworthy difference”. “With [ShiSha] the partner was standing next to me and with [Vishnu] we were standing at the same position. However, the interaction and task solving were equal. The interaction developed a kind of language, to solve the tasks quicker.”

5.3.3 Discussion

The UEQ-S overall scores for both ShiSha and Vishnu are good, as classified by the benchmark of Hinderks et al. [Hin18b]. Only 10% of the benchmark data is better. This strongly supports H_A. Both techniques have high usability and are suitable for this collaborative task.

The average puzzle times show a slight learning effect, as users became familiar with the interaction and improved their task-oriented language. There are no significant differences regarding puzzle completion times and RTLX workload. Yet, there is a slight trend toward the ShiSha system. This supports hypothesis H_B. ShiSha provides similar task performance compared to Vishnu. The shared perspective due to the position-guiding provides users with a good understanding of the puzzle.

Lastly, the results partially support H_C. Both techniques receive high scores on all social presence scales and support the collaboration in this expert/teacher to worker/student setup. Users can understand the actions of the other person. However, the social interaction with Vishnu is limited, because the other user is only partially visible. As user commentary indicates, besides occlusion problems with the other's hands, this can even lead to feeling isolated/lonely or as if being watched. As predicted, the feeling of co-presence is significantly higher with ShiSha than with Vishnu. However, the sub-scales attentional allocation, perceived message understanding, and perceived behavioral interdependence and the aggregated overall score are not significantly different. Written feedback and the look-at times indicate that users observe their partner in the CVE while using ShiSha. Workers look at the expert twice as much as the expert at the worker which emphasizes their dependency on the expert's advice and feedback.

Overall, ShiSha provides an easy-to-use and efficient way to solve the presented task. It achieves equal usability and task performance as Vishnu but yields an improvement regarding co-presence and teamwork.

Yet, there are a couple of limitations with the current implementation of ShiSha. The redirection introduces spatial inconsistencies as the other user's avatar position is not equal to her/his actual position and viewpoint. However, the presented results indicate that most users had no problem working with the social redirection and spatial changes of ShiSha. The task performance and perceived message understanding are high and not significantly different from Vishnu. No guesswork is needed as users can rely on that they know what the other can see, i.e., what they can see. Although users are initially aware of the redirection, they might forget about the manipulation of the CVE due to the high immersion of VR. If a user forgets that only the other user's avatar is shifted, she/he might start referencing objects in the scene incorrectly, e.g., "move this towards you'. However, left-right ambiguities

are also common in real-world conversations and users quickly adapt and can resolve potential issues. A solution might be to add a visualization to inform and steadily remind users about the shift, e.g., show a ghost of the real position or use different colors. Also, a task-based coordinate system (such as cardinal directions in section 5.2) helps to reduce the occurrence of spatial misunderstanding. Advanced speech analysis and synthesis might also be able to inject the correct frame of reference into the communication in future applications.
Another limitation is the fidelity of the user's avatar as it does not represent the real person nor support facial expressions or eye-gaze. Especially the worker relies on the feedback of the expert in the presented training/support scenario and the used avatars convey only a limited amount of information. Recent publications[2] have shown realistic facial expressions and eye-gaze for virtual avatars to be on the verge of availability. These can foster the benefits of ShiSha and are likely to increase its social performance even more.

Although there are large differences regarding user representation, both Vishnu and ShiSha show similar performance results and the interaction feels similar to some users. In addition to that, users comment that even the medium-fidelity avatar already offers great value when collaborating with ShiSha. Also, the task-focused communication between users did not change after they switched the technique. The results support the hypothesis of Sousa et al. [Sou19] that users potentially benefit more from a shared perspective rather than avatar fidelity. During the tasks, participants focus more on the puzzle and use task-oriented communication instead of interacting with the other user socially (see also [Bec05, Mon11, Tan10]). This might explain, why the medium-fidelity of the used avatar or even only virtual arms are sufficient for the goal-oriented tasks. A more open conversation, such as the review session in between puzzles, benefits more from the interpersonal communication that ShiSha offers. 46.9% of users stated that social interactions with ShiSha convey the feeling of teamwork and have a positive effect on the collaboration. Vice versa, the absence of the other user's avatar induces feelings of loneliness or surveillance. Also,

---

[2] e.g., https://tech.fb.com/codec-avatars-facebook-reality-labs/

even though users prefer working with a shared perspective, the overlapping avatars may induce confusion of whose hands are whose [Tan10]. ShiSha solves this problem and allows to distinguish between multiple users while allowing efficient teamwork with a shared perspective.

5.4 Modifying User Avatars to Compensate for Short-Term Connection Loss

When collaborating remotely with other users, the network connection is of crucial importance. Smaller packet loss rates do not influence the communication that much because they can be corrected using packet loss concealment methods, e.g., [Dip17]. However, it happens that the connection of a user is completely interrupted for a short time [Bal16], for example as they transition between wifi access points or the network is briefly overloaded. Apart from connection issues, users might be briefly interrupted by other urgent manners they need to attend to, e.g., the doorbell rang, some quick conversation with another person, quick bathroom break. With smaller groups, the default case of handling a short connection loss is to wait on the absent participant, repeat the conversation that was missed, and then continue with the meeting. For larger groups, this might not be possible and the respective user has to extrapolate the missing conversation or ask about what has happened via another communication channel, e.g., ask colleague through private chat.

Social redirection in VR allows to override user's behavior and adjust content's position, rotation, and scale. Besides manipulating space, it is also possible to redirect a user's avatar in time. The presented approach, called Rewind, targets to compensate for short-term connection loss by introducing two systems. First, if a user loses and regains network connection, a playback repeats the missed-out conversation so that the user can continue to participate in the meeting without information loss [Bro02]. Second, all other users should be affected as little as possible by the absent user. Ideally, the conversation continues without any interrupts as the absent user's avatar is controlled by a compensation strategy.

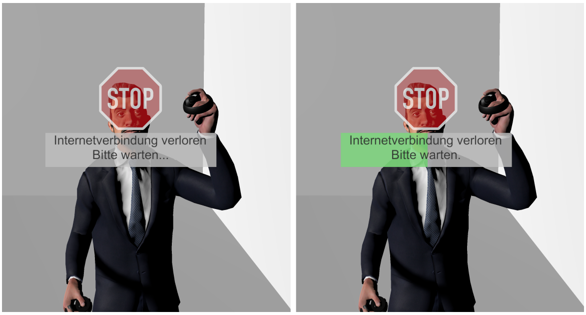


Figure 5.20: The warning message that is displayed when the user has lost the connection (left) and is catching up (right).

As soon as the connection is lost, the respective user needs to be notified about it. For this, a message is displayed that informs the player about the disconnect and the current progress of the playback (see Fig. 5.20). To playback the missed-out information for the user with connection loss, it is required that the server of the CVE logs all necessary data, e.g., speech and movements of users, with their respective timestamps. This is no large overhead because this information is usually present at the server at any time. Once the user reconnects to the CVE, the server resends all[3] information that was gathered while the user was disconnected. The client application then needs to present this information to the user. However, the speed of the playback needs to be increased by a factor S with $S>1$ so that the user can catch up to the current state of the conversation. Also, the server needs to repeatedly send updates to the client, as the conversation continues while the client is catching up. There exists a tradeoff between S being as small as possible so that the user can follow the conversation easily [Aro92] and S being as large as possible to minimize the catch-up time. The total duration of absence is calculated as

$$t_{total} = t_{miss} \cdot \sum_{i=0}^{\infty} \left(\frac{1}{S}\right)^{i} = t_{miss} \cdot \frac{1}{1-\frac{1}{S}} = t_{miss} \cdot \frac{S}{S-1} \tag{5.10}$$

---

[3] The server can perform some filtering to reduce the amount of data, e.g., adjust sampling rate or discretize values to reduce memory size.

where t_{miss} is the duration of the connection loss.

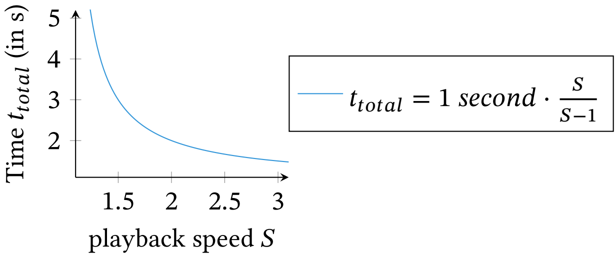


Figure 5.21: Total duration of absence t_{total} with respect for different speed-up factors S for a connection loss duration of $t_{miss} = 1$ second.

Figure 5.21 shows the influence of S on the total duration of absence for a connection loss of 1 second. One might consider strongly increasing S for all parts of the conversation that do not contain speech conversation. This follows the goal to prioritize a shorter catch-up time (minimize t_{total}) rather than an accurate replication of users' movement.

Figure 5.22: Visualization of the mimicry avatar strategy. The opposing avatar takes the same pose as the authority-user.

Because the connection loss of the user is compensated using the playback feature, she/he sends no input to control the avatar. There exist several ways to deal with the avatar.

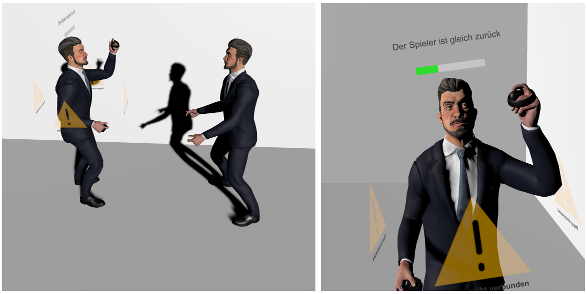


Figure 5.23: Visualization of the explicit avatar strategy. The opposing avatar is surrounded by slowly rotating warning signs. A progress bar above the head shows if the user is disconnected or how long it will take until she/he has caught up.

One strategy to control the avatar is to fully compensate its movements to give the impression that the user is still there. For this, the method of nonverbal mimicry avatar adjustment of Roth et al. [Rot18] was implemented which imitates the movements of the authority-user (see Fig 5.22). The mimicry avatar has the advantage that the conversation is not interrupted but also the disadvantage that the avatar might behave strangely, i.e., when copying pointing gestures from the authority-user or not answering to questions.
A different strategy is explicitly showing that one user has lost the connection and the current state of her/his recovery through a progress bar (see Fig 5.23). With this method, users know when a user is not present and cannot participate in the conversation. Additionally, they can estimate when the user will be back. However, users might stop talking and wait for the absent user due to the explicit visualization.

5.4.1 Evaluation

Setup For the evaluation of the proposed Rewind connection loss compensation, a CVE was implemented using Unity and two HTC Vive Pro HMDs including voice communication. The HMDs were set up in two separate rooms. To guarantee full control over the connection losses, the evaluation was performed using a 1 Gbps Ethernet connection and package loss was simulated by the application. This has the advantage that a connection loss can be determined without relying on the absence of ping messages. Additionally, executing tasks with no connection loss is also possible using this network configuration.

A speed-up factor of S=1.5 was chosen for sections within the playback containing speech and S=∞ for sections without speech (effectively skipping parts where users are not talking).

To evaluate the two Rewind avatar strategies, i.e., *mimicry* and *explicit avatar*, two baseline conditions were implemented as well. The *best-case* baseline contains no connection losses and represents the upper bound for an ideal connection loss compensation. The *worst-case* baseline does not compensate for connection loss at all and users have to manually correct for any issues (as with most current systems). This is the lower bound for the compensation system.

Participants 20 persons (15 male, 5 female) participated as dyads in a within-subject user study. The participants were M=25.4 years old (SD=8.1), with the youngest and oldest users being 18 and 57 years old respectively. 65.0% had no or little experience with VR. 20.0% reported an average and 15.0% a high experience with VR.

Figure 5.24: Outside view of the two users standing at the table with the word cubes on it. One user is grabbing a cube to start the storytelling.

Tasks and Procedures A task was designed that targets to encourage a conversation between the two users. The participants were standing on opposite sites of a table (see Fig. 5.24). In front of them were 40 cubes showing one word each[4]. Half of the cubes contained easy words, e.g., car, dog, house, the other half more complex words, e.g., operating room, company, digital. Users were asked to collaboratively tell a story. One user should pick a cube and start the story based on the given word, e.g., "coffee" ↦ "Alice woke up and made herself a *coffee*". This user should then put the cube on a shelf to her/his left. After that, the other user should pick one of the remaining cubes and continue the story, e.g., "car" ↦ "Alice then got ready for work and got in her *car*". Subsequently, it is the first user's turn again. The goal of the game is to tell a story using as many cubes as possible.

For the user study, the participants filled out an initial questionnaire. After that, users got an introduction to the VR system, the controls, the task, and the meaning of the "STOP"-message and the playback function. 5 sets of 40 cubes were created with no word repetitions. The first set was used for training. Dyads had 7 minutes to learn the controls without any connection loss. After that, users performed the tasks given the remaining 4 sets with one of the conditions: mimicry avatar, explicit avatar, best-case, or worst-case. Each round had a duration of 220 seconds. The conditions were counterbalanced to reduce the effects of training or fatigue. For the three conditions mimicry avatar, explicit avatar, and the worst-case baseline each user experienced 1×5 seconds, 1×4 seconds, and 2×1 second of connection loss. Therefore, the overall duration of connection loss was 10%. The order and pattern of connection losses were varied for each round. After each condition, users took off the HMDs to fill out a questionnaire. In the end, a final questionnaire was presented to dyads.

Measurements As the presented methods target to compensate short-term connection losses, the following hypotheses were formulated for the results of the evaluation:

---

[4] Words were randomly selected from https://www.palabrasaleatorias.com/zufallige-worter.php

H_A Both Rewind avatar strategies increase users' performance compared to the worst-case baseline.

H_B Both Rewind avatar strategies reduce users' workload and increase their social presence compared to the worst-case baseline.

H_C The mimicry avatar adaptation allows to completely mask the connection losses.

To evaluate the language behavior of dyads, their conversations were recorded. It was counted how many cubes were used during the conversation, how often a participant asked her/his partner to repeat herself/himself, how often users repeated themselves, and how often a connection loss was addressed in the conversation. Additionally, users were presented with an RTLX and NMSPM questionnaire for each condition.

5.4.2 Results

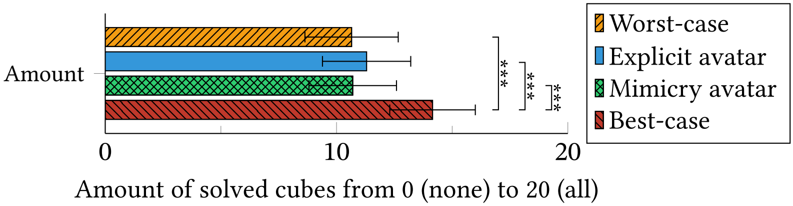


Figure 5.25: Amount of solved cubes per user. Results as average value with 95% confidence intervals and asterisks indicating significant differences.

Quantitative Results Most of all dyads could solve at least half of the cubes with every condition (see Fig. 5.25). Users solved M=14.150 cubes (SD=4.221) with best-case, M=10.650 (SD=4.603) with the worst-case, M=11.300 cubes (SD=4.354) with explicit avatar, and M=10.700 cubes (SD=4.318) with mimicry avatar. All conditions received the maximum value of 20 at least once. Mauchly's test indicates a violation of sphericity ($\chi^2(5)=16.974$, p=.005, Greenhouse-Geisser ϵ=.634). A Greenhouse-Geisser-corrected repeated-measures ANOVA yields a significant difference between the conditions (F(1.903,36.165)=17.004, p<.001, N=20) with a large effect size (partial η^2=.472). Pairwise Bonferroni-corrected comparisons show that the best-case baseline is significantly better than all other conditions (all p≤.001).

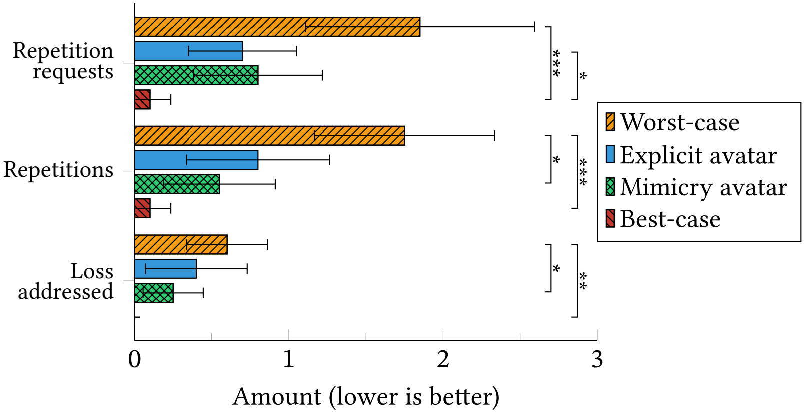


Figure 5.26: Number of times that users asked the other to repeat herself/himself, that a user repeated herself/himself, and that a connection loss was addressed. Results as average value with 95% confidence intervals and asterisks indicating significant differences.

The results of the recordings of the conversation are displayed in Fig. 5.26. Users request the other to repeat herself/himself M=0.100 times (SD=0.308) with the best-case baseline, M=1.850 times (SD=1.694) with worst-case, M=0.700 times (SD=0.801) with explicit avatar, and M=0.800 times (SD=0.951) with mimicry avatar. Mauchly's test indicates a violation of sphericity (χ^2(5)=20.032, p=.001, Greenhouse-Geisser ϵ=.602). A Greenhouse-Geisser-corrected repeated-measures ANOVA yields a significant difference between the conditions (F(1.805,34.289)=8.999, p=.001, N=20) with a large effect size (partial η^2=.321). Pairwise Bonferroni-corrected comparisons show that the best-case baseline is significantly better than worst-case (p=.001) and the avatar conditions (both p≤.041). The number of times that users repeated themselves shows similar results with M=0.100 (SD=0.308) for best-case, M=1.750 (SD=1.333) for worst-case, M=0.800 (SD=1.056) for explicit avatar, and M=0.550 (SD=0.826) for mimicry avatar. Mauchly's test indicates a violation of sphericity ($\chi^2(5)$=12.242, p=.032, Huynh-Feldt ϵ=.880). A Huynh-Feldt-corrected repeated-measures ANOVA yields a significant difference between the conditions (F(2.639,50.146)=10.862, p<.001, N=20) with a large effect size (partial η^2=.364). Pairwise Bonferroni-corrected comparisons show that the best-case baseline is significantly better than worst-case (p<.001) and explicit avatar conditions (p=.041). Also, the mimicry avatar condition is significantly better than the worst-case baseline (p=.028). Users addressed the connection loss only when using a condition that compensates for the connection loss. Therefore, the amounts are M=0.000 (SD=0.000) for best-case, M=0.600 (SD=0.598) for worst-case, M=0.400 (SD=0.754) for explicit avatar, and M=0.250 (SD=0.444) for mimicry avatar. Mauchly's test indicates a violation of sphericity ($\chi^2(5)$=14.618, p=.012, Greenhouse-Geisser ϵ=.683). A Greenhouse-Geisser-corrected repeated-measures ANOVA yields a significant difference between the conditions (F(2.050,38.950)=5.647, p=.007, N=20) with a large effect size (partial η^2=.229). Pairwise Bonferroni-corrected comparisons show that users mention a connection loss significantly less with the best-case baseline than with the worst-case condition (p=.002). Additionally, the mimicry avatar condition is significantly different from worst-case (p=.028).

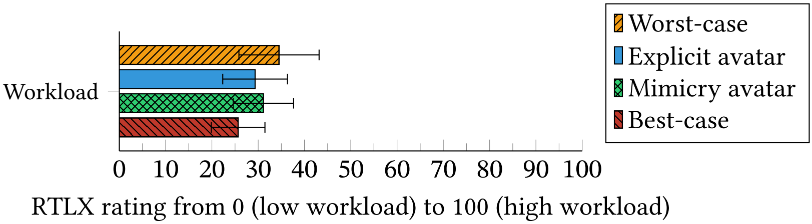


Figure 5.27: RTLX scores. Results as average value with 95% confidence intervals.

Qualitative Results The RTLX workload of users is M=25.667 (SD=13.225) with best-case, M=34.500 (SD=19.760) with worst-case, M=29.333 (SD=15.942) with explicit avatar, and M=31.125 (SD=14.891) with mimicry avatar (see

Fig. 5.27). A Friedman test shows no significant difference between the conditions (χ^2(3)=4.362, p=.225, N=20).

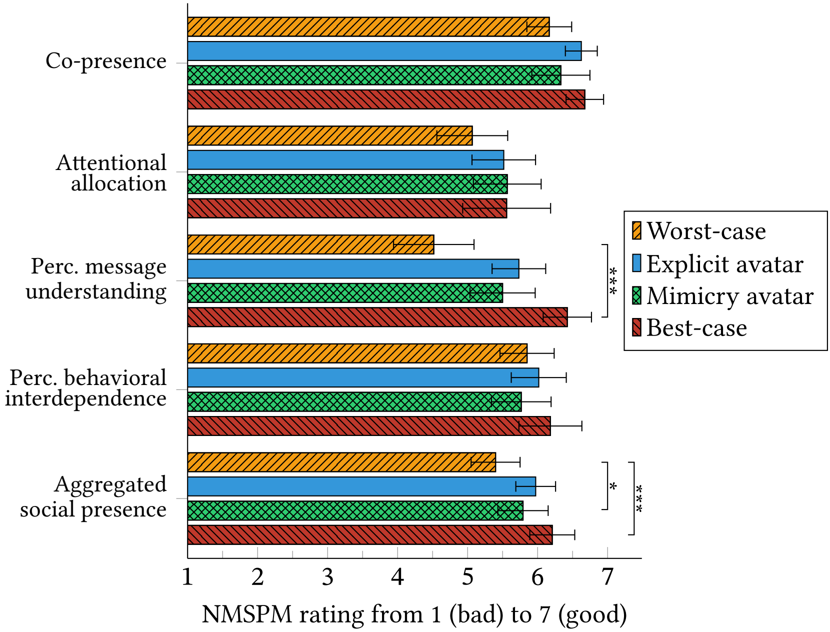


Figure 5.28: NMSPM scores. Results as average value with 95% confidence intervals and asterisks indicating significant differences.

The different scales of the NMSPM questionnaire and an overall aggregated score are shown in Fig. 5.28. Users rated the co-presence as M=6.675 (SD=0.611) for best-case, M=6.167 (SD=0.729) for worst-case, M=6.625 (SD=0.521) for explicit avatar, and M=6.333 (SD=0.947) for mimicry avatar. A Friedman test shows significant differences for co-presence (χ^2(3)=11.743, p=.008, N=20) with a small effect size (Kendall's W=.196). However, Bonferroni-corrected pairwise comparisons show no significant differences. Yet, the difference between the worst-case and best-case conditions is almost significant (p=.051). The attentional allocation of users was perceived as M=5.558 (SD=1.432) for best-case, M=5.067 (SD=1.155) for worst-case, M=5.517 (SD=1.036) for explicit avatar, and M=5.567 (SD=1.105) for mimicry avatar. A Friedman test shows no significant differences for the conditions (χ^2(3)=5.086, p=.166, N=20). The perceived message understanding was rated as M=6.425 (SD=0.788) for best-case, M=4.517 (SD=1.311) for worst-case, M=5.733 (SD=0.874) for explicit avatar, and M=5.500 (SD=1.064) for mimicry avatar. A Friedman test yields a significant difference (χ^2(3)=26.542, p<.001, N=20) with a medium effect size (Kendall's W=.442). Bonferroni-corrected pairwise comparisons show that best-case achieves a better result than worst-case (p<.001). Users rated the perceived behavioral interdependence as M=6.183 (SD=1.027) for best-case, M=5.850 (SD=0.883) for worst-case, M=6.017 (SD=0.897) for explicit avatar, and M=5.767 (SD=0.972) for mimicry avatar. The behavioral interdependence was not significantly influenced by the conditions (χ^2(3)=5.281, p=.152, N=20). Lastly, the systems reach an overall aggregated social presence score of M=6.210 (SD=0.732) for best-case, M=5.400 (SD=0.799) for worst-case, M=5.973 (SD=0.648) for explicit avatar, and M=5.792 (SD=0.820) for mimicry avatar. A Friedman test yields a significant difference (χ^2(3)=20.062, p<.001, N=20) with a medium effect size (Kendall's W=.351). Bonferroni-corrected pairwise comparisons yield significant differences between best-case and worst-case (p<.001) and mimicry avatar and worst-case (p=.024).

Generally, users did not notice a connection loss with the worst-case baseline nor the mimicry avatar conditions. They only became aware of a disconnect, if the user was interrupted mid-sentence.
While using the worst-case baseline, i.e., no connection loss compensation, the freezing of avatar movements was not enough to notice a connection loss. Consequently, most users tried to resolve any issues/ambiguities by asking the other to repeat herself/himself. However, especially the very short interrupts were rarely noticed. This sometimes had larger consequences for the communication. A dyad of participants experienced a one-second disconnect while each telling the other to go first. As there was no playback, both users then waited for the other to start the story. After one minute of silence, they noticed the issue and repeated their negotiation of who may start. Two dyads sometimes ignored the connection losses. They tried to just continue with the story-telling to clear as many cubes as possible.

As with the worst-case baseline, the mimicry avatar adaptations were mostly unnoticed. However, some users recognized that the other avatar was mirroring their movement, e.g., if they were grabbing an object and the other avatar performed the same motion. As a result, some users bridged the duration of the disconnect by performing large gestures or even small dances.
The explicit avatar condition was mostly noticed by the participants. Some users stopped talking as soon as they saw that the other user was not present anymore. They continued as soon as the progress bar indicated that the other user was back. Sometimes they even repeated themselves even though the other user was informed by the playback. Also, the connection loss was addressed by users as they were aware of it.
After the user study, participants expressed the wish for more control over the playback speed factor S. Others perceived the system to be beneficial and asked if a similar system is available for online video conference tools.

5.4.3 Discussion

As indicated by the results and the statements of users, the Rewind playback and avatar adaptations are beneficial to compensate for short-term connection loss. The Rewind system tends to lie between the best-case (i.e., no connection loss) and worst-case (i.e., no compensation for a disconnect) baselines. The task performance measured by the number of solved cubes was not improved. Therefore, H_A cannot be supported. However, the number of repetitions and addressing of the connection loss could be significantly reduced with the mimicry avatar adjustments.

H_B can only be partially supported. Users' workload is not significantly different with the conditions. But, the social presence can be improved significantly with the mimicry avatar adaptation.

It is questionable whether the mimicry avatar adaptation is actually needed to cover up the connection loss of a user. Most users cannot notice that the other user is gone, even if the avatar pose freezes totally as with the worst-case baseline. This indicates a support of H_C. Also, the perceived behavioral interdependence is not increased while using the mimicry avatar manipulations even though the other avatar sometimes exactly mirrors the user's movements. It seems like any user movement, maybe even an idle pose animation, could be used to obscure that a user is not present anymore. This would also alleviate any issues with mirroring pointing and grabbing gestures [Rot18].

The current implementation has some limitations. First of all, the system is a prototype that simulates the connection loss rather than inducing real network issues. The algorithms may need to be adjusted and made more robust for real conditions. Furthermore, the playback speed increase prioritized speech over user movement which might be a wrong assumption in the spatial environment. Also, the pitch of the user's voices increased during the speed-up. A more sophisticated approach might further increase the comprehensibility of the conversation.

The behavior of users while using the explicit avatar compensation highlights that the features of the Rewind system are new to users. Some users seemed to mistrust the system, as they preferred waiting until the user was back and then repeated themselves. The explicit notification seems to interrupt the conversation as it draws attention to the disconnect. However, an explicit highlight of issues could also be beneficial as the worst-case baseline and mimicry avatar adjustments do not allow to recognize when a user loses the connection.

5.5 Conclusion

This chapter presents the concept of PP. The PP modification gives users the possibility to share a common viewpoint by altering spatial relations within the VE. The proposed method allows user-individual adaptations of content through affine transformations such as translation, rotation, and scaling. Resulting world inconsistencies are removed by applying redirection methods in a social context, i.e., through adaptations of users' avatars and tools. The presented implementations offer increased task performance and social presence in comparison to unmodified or baseline methods.

A basic avatar modification adjusts the pointing targets and focus of a user. By rotating the content on a table, two (or even more) users can review data or execute a task presented to them from a mutual perspective. The introduced method transforms indicated positions into the different perspectives of the respective users. Further modification of the user's avatar via rotations of the point of attention and hand orientation results in a mostly undisturbed HHI. A user study showed, that executing tasks in an upright or familiar perspective reduces mental load and increases task performance significantly.

As the basic avatar only changes users' avatars using rotational offsets, it does not allow direct object manipulation. Therefore, an extended avatar modification method, called ShiSha, was developed that is also applicable to full-body avatars. ShiSha is a social redirection technique that allows multi-user collaboration from a shared perspective while at the same time providing a face-to-face interaction by shifting and modifying remote users' avatars. To compensate for the shift, the other avatars are modified to still reach to and look at the correct locations. A user study confirms that the ShiSha modification is easy to use and has an equal performance when compared to a user collaboration with a shared perspective and overlapping avatars. Co-presence scores and user commentary indicate that the face-to-face interaction of ShiSha conveys the feeling of teamwork and allows one to follow the actions of the respective other. The social redirection of ShiSha is especially helpful in the context of remote training or support but is generally applicable for virtual conferences and user guidance.

Especially while working remotely, the quality of a user's connection has a large impact on a conversation. To compensate for short-term interruptions, e.g., through a connection loss or a short turn away due to a distraction, the PP avatar adjustments were extended to consider the course of time. The Rewind system shows a quick playback of the missed conversation to users that were disconnected. During the time the user is absent, mimicry or explicit avatar adaptations can either cover up the connection loss or help to notify users that a person is missing. Rewind increases the perceived message understanding of users compared to an uncompensated disconnect and proves to be beneficial for conversations in VEs.

When implementing collaborative systems and social redirection, one should consider the following insights and design recommendations:

Personal modifications benefit collaboration: The results of the user studies show that the PP technique combined with social redirection avatar adaptations has significant advantages over unchanged VEs that follow the rules of the real world. Spatial inconsistencies due to the modifications can successfully be compensated by the system and users. PP modifications result in equal or better task performance, induce higher feelings of co-presence and teamwork, and increase message understanding when compared to baseline methods. Social redirections can be used for several use cases including increasing teamwork while working with a shared perspective, resolving technical limitations, adjusting virtual content to fit each user's requirements, and altering the behavior of avatars to cover up issues such as a connection loss.

Maintain bodily consistency for other users' avatars: Because following the other user's movements during collaboration is important, behavioral realism should match the appearance of the virtual avatar. As social redirection adjusts the behavior and movements of users, systems should try to minimize noticeable effects during redirection, e.g., unnatural pose or strange movements. Luckily, as users are unaware of how the other user moves in the real world, the redirection does not need to match the real body movements precisely but only present a coherent behavior.

Visualize an avatar for other users: The presented results indicate that although high fidelity user avatars might not be required for task execution, they benefit teamwork and contribute to a natural and comfortable conversation. Feedback of users suggests that observing the movement of the full-body avatar of the other user is an important part of the collaboration. Working together with another avatar induces a feeling of teamwork. Whereas, a missing user avatar leads to issues that the actions of others are not noticeable and cannot be followed as easily. Some users even experience the feeling of loneliness or feel monitored (also see [Smi18]).

Keep personal space distances: As assumed, the requirements for personal space distances from real-world collaborations also translate to VR applications. Methods using overlapping user avatars break personal space distances. Also, overlapping avatar methods pose additional problems as body parts from other users may block one's view. The other user's movements can even be irritating. Social redirection techniques can resolve these issues and position users to maintain a comfortable interaction. As an added benefit, this also introduces the possibility of spatial partitioning which may be helpful in environments with more than two users.

6 Application of a Virtual Map Table

This chapter presents a showcase application that integrates several ideas and insights of previous chapters to display their usefulness for VR systems. This application contains a virtual map table for situation visualization and analysis. As any data that depicts the earth is inherently 3D, the virtual map table represents the spatial data on a virtual globe to provide a better understanding of the terrain. The presented VR system builds upon the GUI of a conventional 2D map display to foster existing functions and familiarity of users with the interface. Different view modes enable detailed or broader inspection of the terrain. Furthermore, VR enables a location-independent multi-user cooperation for quick decision-making.
The content of this chapter is partially based on the following publication [van20].

6.1 Virtual Map Table

The existing web-based technology of a 2D digital map table[1] provides users with tools to visualize and annotate geographical data. The map data is visualized with respect to a date and can therefore change over time. The interface allows the control of the layer-based viewport (e.g., enable/disable layer), views (i.e., bookmarks of viewports, time, and layers), several tools for annotation (e.g., draw lines or polygons), other tools (e.g., measurement), and the timeline (e.g., jump to date or play live time). The application supports a variety of data interfaces, e.g., geospatial or tracking data, and is highly adaptable. Users, e.g., from police, disaster control, fire department, or military, can

---

[1] https://www.iosb.fraunhofer.de/en/projects-and-products/digitaler-lagetisch/

collaboratively assess situations and make decisions using the given application. They can either work independently using a role-based collaboration on multiple distributed devices or they can work synchronously while sharing a large display system. However, as map data occurs naturally within the 3D real world, a 2D map display lacks information. Therefore, the 2D digital map table implements a 3D mode where users can control a virtual camera on a 3D globe. However, even in this mode, the data is visualized on a 2D screen and some information is lost. Also, traditional 2D map print-outs suffer from information loss as they map terrain to contour lines.

Because of the limitations of a 2D map display, a virtual globe application was developed that conveys the map content to the user in real 3D using the stereoscopic visualization of VR. The goal of the virtual map display is to reenable users to use their real-world experiences of terrain e.g., to judge heights and distances even when not physically present on-site. The virtual map table extends the existing range of 2D devices with an immersive data visualization (see Fig. 6.1). However, the 3D map does not target to replace existing 2D maps as a fine-grained understanding of the terrain might not always be necessary or technology is not available, e.g., a user in the field might create ad-hoc maps based on sticks and stones or pen and paper. The application was implemented using Unity and supports current-gen VR HMDs such as HTC Vive Pro, Oculus Rift S, or Valve Index. To display a virtual map within VR, several steps are necessary which are explained in the following paragraphs.

Figure 6.1: The different devices of a software system for a shared situation visualization and analysis.

First, a virtual globe was developed that supports different geospatial data interfaces. The globe is implemented as a mesh-based object. To recreate the shape of the earth accurately, the globe is based on the WGS84 reference ellipsoid[2]. As the shape of the earth differs from the smooth surface of the ellipsoid, elevation data is included to represent terrain changes. This data is stored as a numerical height array and allows to displace the surface of the globe's mesh. Georeferenced imagery is displayed as a texture on the surface of the globe. As with the 2D map display, several textures can be layered on top of each other with varying transparency. The globe is not represented as one large mesh but is composited of several tiles based on a quadtree structure. The tiling enables multiple levels of detail as well as performance optimizations, e.g., culling. Additionally, 3D models, e.g., 3DTiles[3] or glTF[4], can be positioned on or above the globe's surface. All in all, this yields a 3D virtual representation of the earth (compare Google Earth VR[5]).

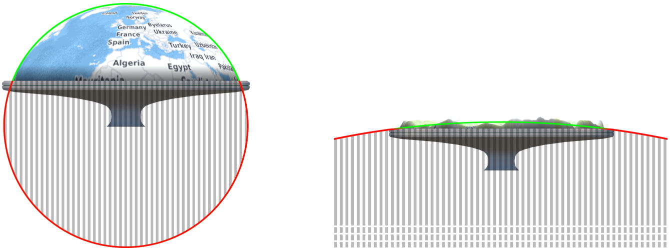


Figure 6.2: The positioning of the 3D globe on a virtual table for a zoomed-out (left) and zoomed-in (right) viewport.

The 2D application allows users to pan and zoom the viewport to display the intended map extent. To allow users to interact with the globe similarly, the decision was made that the globe should be presented on a virtual table. The table has a circular shape as it fits the ellipsoid and allows users to maintain

---

[2] https://epsg.io/4326

[3] https://www.ogc.org/standards/3DTiles/

[4] https://www.khronos.org/gltf/

[5] https://arvr.google.com/earth/

the same distance to the table's center regardless of their positioning around the table. However, other shapes based on a rectangle or torus are possible as the table is only virtual. See Fig. 6.2 for two different examples of map extends. The two images show that the curvature of the earth is not really visible if users work on a smaller geographic extend but only comes apparent as users zoom out. Also, to fit on the table, the globe needs to be positioned and truncated according to the current viewport. As the globe is not uniform but more shaped like a flattened sphere, the positioning of the globe is not straight-forward. To correctly position the globe at the table, the extend of the viewport in the local coordinate system of the globe is mapped on the target extent, i.e., the table's surface, in the world coordinate system of Unity. Furthermore, as the globe can become very large, the float-precision of the engine is not precise enough and the positioning of Unity cannot be used. Therefore, all position calculations are performed using double-precision and only applied to the meshes at the end in the vertex shader on the graphics processing unit (GPU). Because of the positioning on the table, only a section of the globe needs to be visualized. Tiles that are not on the table and those that are not viewed by the user do not need to be loaded or displayed. As tiles' positions are not correct within the Unity engine, the engine's occlusion culling[6] and ray-casting are extended respectively. Furthermore, a level of detail system was added that enables or disables tiles depending on their size within the user's field of view, i.e., based on their screen space error. However, the implemented culling system only removes entire tiles. It is still possible that certain parts of a tile extend over the edge of the table. Therefore, the globe's tiles are additionally culled on a per-pixel basis by discarding the respective pixels in the GPU's fragment shader.

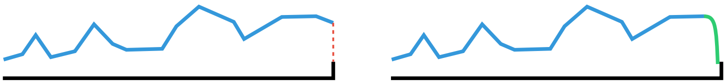


Figure 6.3: The adjustment at the border of the table for a smooth edge. The red line indicates the gap that is present before the adjustment (left). The green line indicates the adjustment of the globe to yield a smoother border (right).

---

[6] https://docs.unity3d.com/Manual/OcclusionCulling.html

As mentioned, the table has a circular shape as this fits the round shape of a globe positioned on it fairly well. However, mountains or buildings may leave holes at the border of the table (see Fig. 6.3). This would allow users to look under the table which might be irritating. To get a closed border at the edge of the table, the mesh of the globe is pulled down. If necessary, meshes that overlap the table's border are subdivided to assure a smooth gradient. As this artificial data does not represent real-world data, a color gradient is added that fades out the globe's texture with increasing divergence.

Figure 6.4: Two users collaborating in the VE from a side (left) and bird-eye (right) view.

One advantage of the VR application is that users can meet within a virtual room without being physically present at the same location. Based on the previous work, full-body avatars were integrated into the virtual map table system (see Fig. 6.4). The ray-casting interaction allows users to precisely indicate a location. Additionally, spatial audio enables users to relate the voices of others to their respective avatar positions. The face-to-face interaction provides a natural collaboration and creates awareness of the other's actions. Furthermore, it does not suffer from the disconnection of users and content within a video conference, i.e., separate video of user and map with a mouse cursor.

6.2 Integration of a Familiar Interface Using the Hybrid Interaction System

The virtual map table is an extension of a 2D map application to 3D displays. However, the VR application does not only support the same features but builds upon the existing GUI. Using the HIS (see section 3.2), the original 2D application is provided as a texture on a plane within the virtual world. Users can interact with the integrated application using the mouse-based ray-casting interaction.

Furthermore, the 2D map display of the desktop application is replaced by the 3D virtual globe as the spatial data is particularly suitable for the immersive visualization. Based on the HIS, a mouse-based interaction using a virtual laser pointer lets users navigate the virtual map using a pan or zoom interaction. Users can pan the current extent by *grabbing* the globe with the laser pointer while holding the left-click button, i.e., trigger. Users can also zoom in or out to change the size of the extent. This can be achieved by using the scroll button, i.e., joystick or touchpad, on a VR controller that is pointing at the map. The map zooms at the location of the laser pointer's intersection with the map. Besides zooming, a horizontal scroll interaction can also be used to rotate the table. A two-handed zoom is also available, e.g., when the controller does not have a joystick. The map zooms out if two grabbing hands are moved closer together. It zooms in if they are moved further apart. In contrast to the similar pinch gesture on touch devices, the map can zoom at the position that the user is looking at (approximated by the HMD's forward direction [Sti02]) which feels natural. Additionally, users can annotate the virtual map by drawing freehand, line string, circle, or polygon shapes. Also, users can place images or text on the map. For this, users point on the map and select different locations using the left-click button. The user finishes the drawing by performing a double left-click.

However, to activate the drawing functions, the VR application needs a UI. As the structure of the 2D drawing menu is quite complex and well suited for a 2D representation [Dar05], it makes sense to just reuse the existing GUI within the VE. Users are already familiar with the layout and function of the content and, e.g., already know how to activate the drawing mode or how to control the different settings such as draw shape, color, width, etc. To connect the 2D web-based and 3D VR-based applications, the message interface of the HIS was implemented. The two applications exchange information such as the current extent of the globe, the current time, the activated layers, and the currently active drawing tool and settings. The two web and VR applications react to these messages and change the displayed content or activate certain functions accordingly. For example, when a user selects the line drawing tool from the 2D window, the respective functionality is activated within the VE which then allows the user to draw directly on the map. After the user finishes the drawing, the geometry of the line is sent back to the web application and further processed there, e.g., send to persistent data storage. Furthermore, the message interface allows the VR application to inherit the functionality of the web application. For example, all web clients are connected via a central server and can send and receive client-to-client messages. The *synchronize viewport* feature of the web application, which synchronizes the map display across all participants (if enabled), now also extends to the VR map table. As a VR user pans her/his viewport, the information is sent to the web page, to the central server, to all other web pages, and finally to other VR users. Therefore, multi-user synchronization does not need to be implemented for the virtual map table within VR.

A first prototype integrated the whole web application as a giant screen on the wall and gave users access to all existing functionality. However, the wall display was cumbersome to read and control as users needed to move around to reliably interact with the window or had to adjust the window's position as they moved around the scene. Another approach could have been to use a handheld window, but some settings are quite small and the window would become too large. Because of this, instead of just displaying the 2D window as-is, the existing UI was separated into different parts using the proposed VR-mode. The VR-mode is activated via the message interface at the start of the VR application. The 2D menu is adjusted to present all menus of the interface in a flattened view (see Fig. 6.5). Furthermore, a circular main menu is added that replaces the original menu controls. This new main menu is then attached to the hand of the VR user (see Fig. 6.6). The web application sends updates to the VR application when a user activates or closes a certain menu. The respective part of the web application is then cut-out from the VR-mode window and displayed in the VE. To extend the functionality of the cut-out windows that are initially attached to the user's hands, a grabbing function was integrated that lets users detach and position the sub-windows at will. Floating windows can be reattached to the hand by selecting the respective menu item from the main window, or closed by dropping the window into the main circle. This allows users to arrange different sub-windows depending on their current requirements.

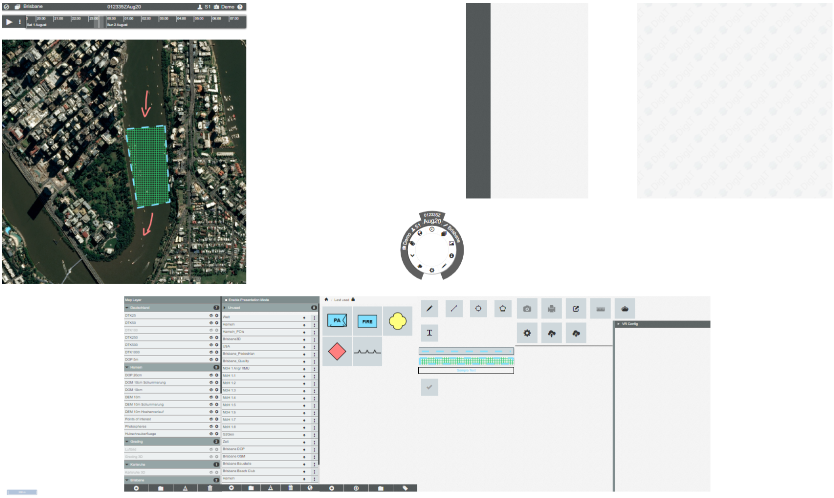


Figure 6.5: The flattened UI of the 2D application in VR-mode

Reusing the existing interface and integrating it within the VR world using the HIS has several advantages. The familiar interface and controls allow users to directly work with the VR application without having to learn new menu structures. Furthermore, the efforts for development are drastically reduced (especially regarding time and errors) as the existing application is stable and its functionality is not replicated into another code-base. Changes, e.g., new features or bug fixes, in the 2D application are also automatically

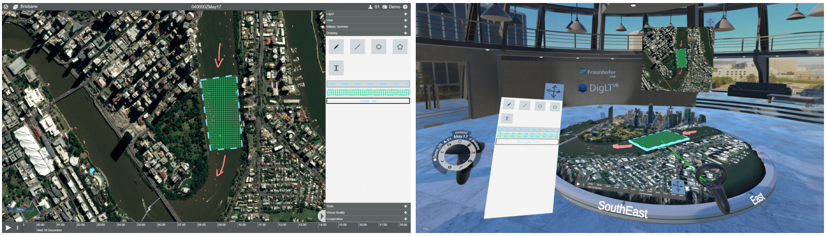


Figure 6.6: The 2D interface (left) that is familiar to users is reused within the VR application using the HIS and the VR-mode (right).

transferred into the VR application. Lastly, the transfer of features to the VE can be done piece by piece as all functionality is available from the start.

6.3 Virtual Site Inspection

Figure 6.7: Virtual site inspection from a rooftop (left) and a construction site (right) using the pedestrian mode.

Viewing geospatial data on a 2D or 3D display gives users a good impression of the coarse surroundings and the area of interest. The quality of the data plays an important role when working with the virtual globe. However, standing around the virtual map table also gives a different impression than actually being on-site. Distances, the width of a passage, or the height of a building can be easily assessed when standing in front of the respective object of interest. To mimic the real-world experience of being in situ, a virtual site inspection, or *pedestrian mode*, was implemented. The pedestrian mode allows users to walk, teleport, or fly through a virtual world thereby enabling users to get a realistic idea of the environment for the assessment of terrain. Figure 6.7 shows the user standing in different locations within a virtual replication of a city that was reconstructed from drone imagery. To activate the pedestrian mode, users point at a location on the table and perform the teleport interaction. After that, the virtual globe is scaled up to represent the content at a scale of 1:1. The user is then placed at the pointed-at coordinate. The globe is scaled up rather than the player is scaled down because of technical limitations. It is not possible to shrink the user to experience the virtual world at the correct size. Furthermore, the shape of the table would be too restrictive to explore the virtual globe. Therefore, the virtual table is removed and the table's radius is increased drastically to get a larger range of sight. As users start moving around the virtual site using the locomotion techniques, they are actually not moving but the globe rotates under them. This has the advantage that users can explore in any direction without hitting any invisible border. Additionally, users always stay approximately at the center of the coordinate system which avoids floating-point precision issues.

While working with multiple users inside the virtual room, each user might follow a different goal. As described, the map extend that is displayed on the globe can be synchronized between all users. However, some users might want to disengage the synchronization to fulfill their respective tasks. To compensate for the differently positioned, rotated, or scaled content on the virtual map tables of users, PP modifications are used (see section 5.1). The pointing target and avatars of users are adjusted to compensate for the different map extends. To convert between the different worlds, the positions are transformed into the local coordinate system of the globe and then synchronized across the connected VEs. As presented in the previous chapter, the PP modifications allow several users to work with e.g., a north-oriented map while standing side-by-side, facilitating the advantages of a shared perspective and face-to-face communication. For this, users simply rotate the table to fit their desired perspective. Similarly, users can work with differently scaled content, e.g., while using different table view modes. Figure 6.8 shows two users working together while one user is viewing the globe from a table perspective and another user is viewing the globe using the pedestrian mode. As in the work of Piumsomboon et al. [Piu18], the social redirection, in this case, is used to overcome the technical limitations of the system as users see the globe at a different scale.

Figure 6.8: The PP modifications let multiple users experience the virtual content at different scales, e.g., from a virtual site inspection (left) and table (right) view.

6.4 Conclusion

The target of this chapter is to show the practical applicability of some of the ideas and features that were presented in earlier chapters. The presented virtual map table enables users to experience a virtual replication of the earth for situation visualization and analysis. Multiple users can collaborate at co-located and remote locations and meet within the virtual room.

The integration of an existing 2D web application using the HIS yielded a fully-featured VR application from the start of development. Further extensions allowed to increase the usability of the presented VR application step by step. The spatial content of the web page was reimplemented within the VE and synchronized with the base application using a message interface. Moreover, the web application was extended with a VR-mode to enable a richer interaction within VR. The implementation of the HIS was fairly quick as major parts of the 2D interface and functionality could be reused.

Additionally, different view modes were implemented that allow users to explore the virtual map on varying scales and rotations while employing different perspectives and modes of locomotion. The integrated PP modifications maintain the spatial relationship of users and content to allow multiple users to collaborate with an individual view of the map. The adaptations allow the synchronization of several virtual globes with map extents that are not aligned between users, e.g., to present a shared perspective or to enable different view modes.

7 Conclusion

This thesis presents three key aspects that increase the productivity and usability of VEs: Integration of 2D hardware and software into VEs, enabling co-located and remote collaboration, and spatial and temporal modifications that enhance multi-user collaboration by providing personalized perspectives. Parts of the work were integrated into an application to showcase their practical applicability.
This thesis shows that VR experiences can be based on real-world skills and knowledge that are transferred into the VE. However, further modifications that use the unrestricted capabilities of virtual worlds allow to enhance this experience beyond real-world limitations.

Commercial off-the-shelf 2D input hardware and software are integrated into VEs to increase the input efficiency and functional range of VR systems. By providing users with virtual proxies of commonly used input devices, i.e., keyboard and tablet, users achieve input accuracies comparable to common VR input methods such as a magic wand or gesture controls due to the haptic feedback of the real-world device and the visual feedback of the virtual model. User's learning efforts are decreased as they can build upon their preexisting real-world skills. As the virtual proxy and the user's hands are both virtual representations of real-world objects, high tracking accuracy is decisive for a precise interaction. Because the devices and hands are purely virtual, diverse adaptations can be employed to alter the input experience, e.g., transparent devices/hands or context-sensitive button labels. This possibility of further enhancements of the virtual proxies increases their potential for VR input even further.
As current VR software is highly specialized and custom-tailored to specific use cases and hardware, the available functionality within VEs is low. Easy

access to common tools, e.g., web browser or office, is missing. By integrating existing 2D applications into a VE, the issue is resolved. A mouse-based interaction method allows a uniform transition between the 2D and 3D content. Additionally, suitable parts, e.g., spatial content, can be transferred to VR to foster the benefits of a stereoscopic visualization. A message interface between the 2D and VR application ensures state synchronization and allows to exchange commands, e.g., activate a VR tool based on a 2D menu item selection. Also, a VR-mode within a 2D application allows to implement an adaptive menu interaction within VR, e.g., automatic sub-window activation and positioning or pop-up windows.

To enable co-located collaboration of multiple VR systems, their tracking spaces need to be aligned, i.e., calibrated. However, because currently available systems with 6DOF tracking employ diverse sensors, algorithms, and operating systems, a uniform calibration method is not available. Therefore, an algorithm is presented that allows the calibration of a pair of any 6DOF devices based on the minimum available data, i.e., position and angular velocity. The presented algorithm enables precise collaboration of two or more 6DOF devices and has no requirements for tracking sensors and time synchronization. Also, it can be executed on a remote machine for low-power devices.
As users start working in a shared physical space, they might collide with each other. A full-body avatar allows users to reliably reduce collisions and maintains co-presence with other users.
As users start working with virtual avatars, the other user does not necessarily need to be physically present. The immersive visualization through an HMD and an audio headset tunes out reality and blocks out other co-located humans. Because of this, the physical location of users becomes less important and co-located and remote collaboration show little differences.

VR applications can replicate real-world experiences, e.g., to allow realistic training. However, virtual worlds can be further enhanced to overcome real-world limitations to increase productivity, ease-of-use, teamwork, or to decrease exhaustion.

Using spatial transformations, each user can be provided with a personal perspective that is based on her/his requirements or needs. For example, content can be rotated upright or scaled up to be easily readable. As these individual modifications lead to inconsistencies between users, further modifications for the virtual avatars are introduced. The personal modifications increase task performance and the feeling of teamwork.
Furthermore, temporal modifications allow to bridge short-term breaks in conversations, e.g., because of a connection loss. By masking the connection loss of a user and presenting her/him with a replay of the missed conversation, co-presence and message understanding can be improved significantly.

Future work, beyond the scope of this thesis, may address two aspects: First, additional improvements or extensions of this work may increase the productivity of the presented systems even further. The performance of the virtual proxy interaction may be increased using more precise tracking. Also, other input devices, e.g., a gamepad, might be included. Automatic content or usage analysis may increase the ease of implementing the proposed VR-mode to enable even easier integrations of existing 2D software. A personalized avatar with facial expressions and eye gaze for additional feedback cues will probably enable more comfortable cooperation. Further work may identify and support more use cases for the presented personal perspective adaptations, e.g., enabling accessible and transparent collaboration of users with different needs. The behavioral realism of the distance-based social redirection of avatars might also be further improved using machine learning, e.g., using synthesized movements [Sta19].
Second, a long-term goal of technology might be to blur the line between devices such as desktop PCs, smartphones, VR, and AR systems (compare Milgram's reality-virtuality continuum [Mil95]). Currently, each device has its own operating system, menu structure, and control scheme. It would surely be beneficial if a seamless transition between the different devices would be supported. Several manufacturers work towards consistent file access, e.g., through cloud storage. But, an easy switch between devices needs more than that. As devices themselves might become more and more functionally equal, users might start to decide which device to use based upon availability and

task requirements. To foster and maintain multi-user collaboration, the software will need to be platform-independent and transparent. Ultimately, this would allow users to choose the device they want and work together with others regardless of hardware capabilities.

Own Publications

This section contains a complete list of own publications. The publications [1], [2], [3], [4] address the transfer of real-world hardware and software into virtual worlds to increase performance and to extend the functional range of a virtual reality application. Publications [5], [6], [7] present foundations of co-located and remote collaboration. The publications [8], [9] show how separated virtual worlds can be leveraged to increase the usability and social presence during task execution. The publication [10] shows the application of a virtual map table. Lastly, the publications [11], [12], [13], [14] can only be considered relevant to the presented work in a broader sense.

[1] Hoppe, Adrian H.; Otto, Leonard; van de Camp, Florian; Stiefelhagen, Rainer and Unmüßig, Gabriel: “qVRty: Virtual Keyboard with a Haptic, Real-World Representation”. In: *HCI International 2018 – Posters’ Extended Abstracts.* Cham: Springer International Publishing, 2018, pp. 266–272. doi: 10.1007/978-3-319-92279-9_36.

[2] Hoppe, Adrian H.; Marek, Felix; van de Camp, Florian and Stiefelhagen, Rainer: “VirtualTablet: Extending Movable Surfaces with Touch Interaction”. In: *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR).* Mar. 2019, pp. 980–981. doi: 10.1109/VR.2019.8797993.

[3] Hoppe, Adrian H.; Marek, Felix; van de Camp, Florian and Stiefelhagen, Rainer: “Extending Movable Surfaces with Touch Interaction Using the VirtualTablet : An Extended View”. In: *Advances in Science, Technology and Engineering Systems Journal* 5.2 (2020), pp. 328–337. doi: 10.25046/aj050243.

[4] Hoppe, Adrian H.; van de Camp, Florian and Stiefelhagen, Rainer: "Enabling Interaction with Arbitrary 2D Applications in Virtual Environments". In: *HCI International 2020 - Posters.* Cham: Springer International Publishing, 2020, pp. 30–36. doi: 10.1007/978-3-030-50729-9_4.

[5] Hoppe, Adrian H.; Kaucher, Leon; van de Camp, Florian and Stiefelhagen, Rainer: "Calibration of Diverse Tracking Systems to Enable Local Collaborative Mixed Reality Applications". In: *Virtual, Augmented and Mixed Reality. Design and Interaction.* Cham: Springer International Publishing, 2020, pp. 63–80. doi: 10.1007/978-3-030-49695-1_5.

[6] Hoppe, Adrian H.; Reeb, Roland; van de Camp, Florian and Stiefelhagen, Rainer: "Capability for Collision Avoidance of Different User Avatars in Virtual Reality". In: *HCI International 2018 – Posters' Extended Abstracts.* Cham: Springer International Publishing, 2018, pp. 273–279. doi: 10.1007/978-3-319-92279-9_37.

[7] Hoppe, Adrian H.; Reeb, Roland; van de Camp, Florian and Stiefelhagen, Rainer: "Interaction of Distant and Local Users in a Collaborative Virtual Environment". In: *Virtual, Augmented and Mixed Reality: Interaction, Navigation, Visualization, Embodiment, and Simulation.* Cham: Springer International Publishing, 2018, pp. 328–337. doi: 10.1007/978-3-319-91581-4_24.

[8] Hoppe, Adrian H.; van de Camp, Florian and Stiefelhagen, Rainer: "Personal Perspective: Using Modified World Views to Overcome Real-Life Limitations in Virtual Reality". In: *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR).* 2018, pp. 577–578. doi: 10.1109/VR.2018.8446311.

[9] Hoppe, Adrian H.; van de Camp, Florian and Stiefelhagen, Rainer: "ShiSha: Enabling Shared Perspective With Face-to-Face Collaboration Using Redirected Avatars in Virtual Reality". In: *Proc. ACM Hum.-Comput. Interact.* 4.CSCW 3 (Jan. 2021). doi: 10.1145/3432950.

[10] van de Camp, Florian; Hoppe, Adrian H.; Unmüßig, Gabriel and Peinsipp-Byma, Elisabeth: “Cooperative and location-independent terrain assessment for deployment planning using a 3D mixed reality environment”. In: *Virtual, Augmented, and Mixed Reality (XR) Technology for Multi-Domain Operations.* Vol. 11426. International Society for Optics and Photonics. SPIE, 2020, pp. 1–8. doi: 10.1117/12.2558593.

[11] Hoppe, Adrian H.; van de Camp, Florian and Stiefelhagen, Rainer: “Interaction with Three Dimensional Objects on Diverse Input and Output Devices: A Survey”. In: *HCI International 2017 – Posters’ Extended Abstracts.* Cham: Springer International Publishing, 2017, pp. 130–139. doi: 10.1007/978-3-319-58750-9_18.

[12] Hoppe, Adrian H.; Anken, Julia K.; Schwarz, Thorsten; Stiefelhagen, Rainer and van de Camp, Florian: “CLEVR: A Customizable Interactive Learning Environment for Users with Low Vision in Virtual Reality”. In: *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility.* ASSETS ’20. Virtual Event, Greece: Association for Computing Machinery, 2020. doi: 10.1145/3373625.3418009.

[13] Hoppe, Adrian H.; Westerkamp, Kai; Maier, Sebastian; van de Camp, Florian and Stiefelhagen, Rainer: “Multi-user Collaboration on Complex Data in Virtual and Augmented Reality”. In: *HCI International 2018 – Posters’ Extended Abstracts.* Cham: Springer International Publishing, 2018, pp. 258–265. doi: 10.1007/978-3-319-92279-9_35.

[14] Maier, Sebastian; Gostner, Till; van de Camp, Florian and Hoppe, Adrian H.: “Outdoor AR-Application for the Digital Map Table”. In: *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLIV-3/W1-2020 (2020), pp. 93–98. doi: 10.5194/isprs-archives-XLIV-3-W1-2020-93-2020.

Bibliography

[Abd19] Abdi Oskouie, Mina and Boulanger, Pierre: “Using Proxy Haptic for a Pointing Task in the Virtual World: A Usability Study”. In: *Augmented Reality, Virtual Reality, and Computer Graphics.* Ed. by De Paolis, Lucio Tommaso and Bourdot, Patrick. Cham: Springer International Publishing, 2019, pp. 292–299 (cit. on p. 18).

[Abt18] Abtahi, Parastoo and Follmer, Sean: “Visuo-Haptic Illusions for Improving the Perceived Performance of Shape Displays”. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.* ACM. 2018, p. 150 (cit. on p. 31).

[Ahl06] Ahlström, David; Hitz, Martin and Leitner, Gerhard: “An Evaluation of Sticky and Force Enhanced Targets in Multi Target Situations”. In: *Proceedings of the 4th Nordic Conference on Human-Computer Interaction: Changing Roles.* NordiCHI ’06. Oslo, Norway: Association for Computing Machinery, 2006, pp. 58–67. doi: 10.1145/1182475.1182482 (cit. on p. 20).

[Aki03] Akiduki, Hironori; Nishiike, Suetaka; Watanabe, Hiroshi; Matsuoka, Katsunori; Kubo, Takeshi and Takeda, Noriaki: “Visual-vestibular conflict induced by virtual reality in humans”. In: *Neuroscience letters* 340.3 (2003), pp. 197–200 (cit. on p. 12).

[All01] Allison, Robert S; Harris, Laurence R; Jenkin, Michael; Jasiobedzka, Urszula and Zacher, James E: “Tolerance of temporal delay in virtual environments”. In: *Virtual Reality, 2001. Proceedings. IEEE.* IEEE. 2001, pp. 247–254 (cit. on p. 11).

[And06] Andujar, C.; Fairen, M. and Argelaguet, F.: “A Cost-effective Approach for Developing Application-control GUIs for Virtual Environments”. In: *3D User Interfaces (3DUI’06)*. 2006, pp. 45–52. doi: 10.1109/VR.2006.6 (cit. on p. 21).

[And07a] Andujar, C. and Argelaguet, F.: “Virtual Pads: Decoupling Motor Space and Visual Space for Flexible Manipulation of 2D Windows within VEs”. In: *2007 IEEE Symposium on 3D User Interfaces*. Mar. 2007. doi: 10.1109/3DUI.2007.340781 (cit. on p. 20).

[And07b] Andujar, Carlos and Argelaguet, Ferran: “Anisomorphic raycasting manipulation for interacting with 2D GUIs”. In: *Computers & Graphics* 31.1 (2007), pp. 15–25. doi: https://doi.org/10.1016/j.cag.2006.09.003. url: http://www.sciencedirect.com/science/article/pii/S0097849306001816 (cit. on p. 20).

[Ang95a] Angus, I. G. and Sowizral, H. A.: “VRMosaic: WEB access from within a virtual environment”. In: *Proceedings of Visualization 1995 Conference*. 1995, pp. 59–64. doi: 10.1109/INFVIS.1995.528687 (cit. on p. 19).

[Ang95b] Angus, Ian G. and Sowizral, Henry A.: “Embedding the 2D interaction metaphor in a real 3D virtual environment”. In: *Stereoscopic Displays and Virtual Reality Systems II*. Ed. by Fisher, Scott S.; Merritt, John O. and Bolas, Mark T. Vol. 2409. International Society for Optics and Photonics. SPIE, 1995, pp. 282–293. doi: 10.1117/12.205875. url: https://doi.org/10.1117/12.205875 (cit. on pp. 19, 20).

[Ann91] Annett, Marian: “Reading upside down and mirror text in groups differing for right minus left hand skill”. In: *European Journal of Cognitive Psychology* 3.4 (1991), pp. 363–377 (cit. on pp. 28, 130).

[Ant16] Anthes, C.; García-Hernández, R. J.; Wiedemann, M. and Kranzlmüller, D.: “State of the art of virtual reality technology”. In: *2016 IEEE Aerospace Conference*. 2016, pp. 1–19. doi: 10.1109/AERO.2016.7500674 (cit. on p. 2).

[Ant18] Anton, David; Kurillo, Gregorij and Bajcsy, Ruzena: "User Experience and Interaction Performance in 2D/3D Telecollaboration". In: *Future Generation Computer Systems* 82 (2018), pp. 77–88 (cit. on p. 138).

[Ara16] Araujo, Bruno; Jota, Ricardo; Perumal, Varun; Yao, Jia Xian; Singh, Karan and Wigdor, Daniel: "Snake Charmer: Physically Enabling Virtual Objects". In: *Proceedings of the TEI '16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction.* TEI '16. Eindhoven, Netherlands: Association for Computing Machinery, 2016, pp. 218–226. doi: 10.1145/2839462.2839484 (cit. on p. 16).

[Aro18] Arora, Rahul; Habib Kazi, Rubaiat; Grossman, Tovi; Fitzmaurice, George and Singh, Karan: "SymbiosisSketch: Combining 2D & 3D Sketching for Designing Detailed 3D Objects in Situ". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.* New York, NY, USA: Association for Computing Machinery, 2018, pp. 1–15. url: https://doi.org/10.1145/3173574.3173759 (cit. on p. 15).

[Aro92] Arons, Barry: "Techniques, Perception, and Applications of Time-Compressed Speech". In: 1992 (cit. on p. 150).

[Art98] Arthur, Kevin; Preston, Timothy; Taylor, Russell; Brooks, Frederick; Whitton, Mary and Wright, William: "Designing and Building the PIT: A Head-Tracked Stereo Workspace for Two Users". In: *2nd International Immersive Projection Technology Workshop.* Chapel Hill, NC, USA: University of North Carolina at Chapel Hill, 1998, pp. 11–12 (cit. on p. 21).

[Aru87] Arun, K. S.; Huang, T. S. and Blostein, S. D.: "Least-Squares Fitting of Two 3-D Point Sets". In: *IEEE Trans. Pattern Anal. Mach. Intell.* 9.5 (May 1987), pp. 698–700. doi: 10.1109/TPAMI.1987.4767965 (cit. on p. 85).

[Azm16] Azmandian, Mahdi; Hancock, Mark; Benko, Hrvoje; Ofek, Eyal and Wilson, Andrew D: "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.* ACM. 2016, pp. 1968–1979 (cit. on p. 31).

[Azm17] Azmandian, M.; Grechkin, T. and Rosenberg, E. S.: “An evaluation of strategies for two-user redirected walking in shared physical spaces”. In: *2017 IEEE Virtual Reality (VR).* Mar. 2017, pp. 91–98. doi: 10.1109/VR.2017.7892235 (cit. on p. 25).

[Azu97] Azuma, Ronald T.: “A Survey of Augmented Reality”. In: *Presence: Teleoperators and Virtual Environments* 6.4 (1997), pp. 355–385. doi: 10.1162/pres.1997.6.4.355 (cit. on p. 10).

[Bab18] Babic, Teo; Reiterer, Harald and Haller, Michael: “Pocket6: A 6DoF Controller Based On A Simple Smartphone Application”. In: *Proceedings of the Symposium on Spatial User Interaction.* SUI ’18. Berlin, Germany: Association for Computing Machinery, 2018, pp. 2–10. doi: 10.1145/3267782.3267785. url: https://doi.org/10.1145/3267782.3267785 (cit. on p. 15).

[Bac13] Bachmann, E. R.; Holm, J.; Zmuda, M. A. and Hodgson, E.: “Collision prediction and prevention in a simultaneous two-user immersive virtual environment”. In: *2013 IEEE Virtual Reality (VR).* 2013, pp. 89–90. doi: 10.1109/VR.2013.6549377 (cit. on p. 25).

[Bac19] Bachmann, E. R.; Hodgson, E.; Hoffbauer, C. and Messinger, J.: “Multi-User Redirected Walking and Resetting Using Artificial Potential Fields”. In: *IEEE Transactions on Visualization and Computer Graphics* 25.5 (2019), pp. 2022–2031. doi: 10.1109/TVCG.2019.2898764 (cit. on p. 25).

[Bai03] Bailenson, Jeremy N.; Blascovich, Jim; Beall, Andrew C. and Loomis, Jack M.: “Interpersonal Distance in Immersive Virtual Environments”. In: *Personality and Social Psychology Bulletin* 29.7 (2003). PMID: 15018671, pp. 819–833. doi: 10.1177/0146167203029007002. url: https://doi.org/10.1177/0146167203029007002 (cit. on p. 30).

[Bai05] Bailenson, Jeremy N; Swinth, Kim; Hoyt, Crystal; Persky, Susan; Dimov, Alex and Blascovich, Jim: "The Independent and Interactive Effects of Embodied-Agent Appearance and Behavior on Self-Report, Cognitive, and Behavioral Markers of Copresence in Immersive Virtual Environments". In: *Presence: Teleoperators & Virtual Environments* 14.4 (2005), pp. 379–393 (cit. on pp. 32, 135).

[Bai17] Bai, Huidong; Gao, Lei and Billinghurst, Mark: "6DoF Input for Hololens Using Vive Controller". In: *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications on - SA '17.* SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications. Bangkok, Thailand: ACM Press, 2017, pp. 1–1. doi: 10.1145/3132787.3132814 (cit. on p. 23).

[Baj12] Bajer, Bartosz; MacKenzie, I Scott and Baljko, Melanie: "Huffman base-4 text entry glove (H4 TEG)". In: *Wearable Computers (ISWC), 2012 16th International Symposium on.* IEEE. 2012, pp. 41–47 (cit. on pp. 13, 39).

[Bal16] Baltrunas, D.; Elmokashfi, A.; Kvalbein, A. and Alay, Ö.: "Investigating packet loss in mobile broadband networks under mobility". In: *2016 IFIP Networking Conference (IFIP Networking) and Workshops.* 2016, pp. 225–233. doi: 10.1109/IFIPNetworking.2016.7497225 (cit. on p. 149).

[Ban13] Banakou, Domna; Groten, Raphaela and Slater, Mel: "Illusory ownership of a virtual child body causes overestimation of object sizes and implicit attitude changes". In: *Proceedings of the National Academy of Sciences* 110.31 (2013), pp. 12846–12851. doi: 10.1073/pnas.1306779110. eprint: https://www.pnas.org/content/110/31/12846.full.pdf. url: https://www.pnas.org/content/110/31/12846 (cit. on p. 13).

[Bar05] Barsoum, Emad and Kuester, Falko: "WebVR: an interactive web browser for virtual environments". In: *Stereoscopic Displays and Virtual Reality Systems XII.* Ed. by Woods, Andrew J.; Merritt, John O.; Bolas, Mark T. and McDowall, Ian E.

Vol. 5664. International Society for Optics and Photonics. SPIE, 2005, pp. 540–547. doi: 10.1117/12.582624 (cit. on p. 21).

[Bat12] Batchelor, Bruce G. and Waltz, Frederick M.: “Morphological Image Processing”. In: *Machine Vision Handbook.* Ed. by Batchelor, Bruce G. London: Springer London, 2012, pp. 801–870. doi: 10.1007/978-1-84996-169-1_19. url: https://doi.org/10.1007/978-1-84996-169-1_19 (cit. on p. 48).

[Bec05] Becker-Beck, Ulrike; Wintermantel, Margret and Borg, Anna: “Principles of Regulating Interaction in Teams Practicing Face-To-Face Communication Versus Teams Practicing Computer-Mediated Communication”. In: *Small Group Research* 36.4 (2005), pp. 499–536. doi: 10.1177/1046496405277182 (cit. on p. 148).

[Bec13] Beck, S.; Kunert, A.; Kulik, A. and Froehlich, B.: “Immersive Group-to-Group Telepresence”. In: *IEEE Transactions on Visualization and Computer Graphics* 19.4 (Apr. 2013), pp. 616–625. doi: 10.1109/TVCG.2013.33 (cit. on pp. 22, 24).

[Bec16] Becker, Torsten; Herrmann, Richard; Sandor, Viktor; Schäfer, Dominik and Wellisch, Ulrich: Stochastische Risikomodellierung und statistische Methoden: Ein anwendungsorientiertes Lehrbuch für Aktuare. Berlin, Heidelberg: Springer Berlin Heidelberg, 2016, pp. 62–65. doi: 10.1007/978-3-662-49407-3 (cit. on p. 294).

[Bel19] Belen, Ryan Anthony J. de; Nguyen, Huyen; Filonik, Daniel; Del Favero, Dennis; Bednarz, Tomasz; 1 Expanded Perception and Interaction Centre (EPICentre), Faculty of Art & Design, University of New South Wales, Sydney, NSW 2021, Australia and 2 CSIRO Data61, Australian Technology Park, Eveleigh, NSW 2015, Astralia: “A Systematic Review of the Current State of Collaborative Mixed Reality Technologies: 2013–2018”. In: *AIMS Electronics and Electrical Engineering* 3.2 (2019), pp. 181–223. doi: 10.3934/ElectrEng.2019.2.181 (cit. on p. 3).

[Ben01] Benford, Steve; Greenhalgh, Chris; Rodden, Tom and Pycock, James: “Collaborative virtual environments”. In: *Communications of the ACM* 44.7 (2001), pp. 79–85 (cit. on p. 3).

[Ben02] Benson, Alan J: “Motion sickness”. In: *Medical Aspects of Harsh Environments 2* (2002), pp. 1048–1084 (cit. on p. 30).

[Ben05] Benko, H.; Ishak, E. W. and Feiner, S.: “Cross-dimensional gestural interaction techniques for hybrid immersive environments”. In: *IEEE Proceedings. VR 2005. Virtual Reality, 2005.* 2005, pp. 209–216. doi: 10.1109/VR.2005.1492776 (cit. on p. 19).

[Ben97] Benford, Steve and Greenhalgh, Chris: “Introducing third party objects into the spatial model of interaction”. In: *Proceedings of the Fifth European Conference on Computer Supported Cooperative Work.* Springer. 1997, pp. 189–204 (cit. on p. 21).

[Bie20] Biener, V.; Schneider, D.; Gesslein, T.; Otte, A.; Kuth, B.; Kristensson, P. O.; Ofek, E.; Pahud, M. and Grubert, J.: “Breaking the Screen: Interaction Across Touchscreen Boundaries in Virtual Reality for Mobile Knowledge Workers”. In: *IEEE Transactions on Visualization and Computer Graphics* 26.12 (2020), pp. 3490–3502. doi: 10.1109/TVCG.2020.3023567 (cit. on p. 17).

[Bil00] Billinghurst, M.; Poupyrev, I.; Kato, H. and May, R.: “Mixing Realities in Shared Space: An Augmented Reality Interface for Collaborative Computing”. In: *2000 IEEE International Conference on Multimedia and Expo. ICME2000. Proceedings. Latest Advances in the Fast Changing World of Multimedia (Cat. No.00TH8532).* 2000 IEEE International Conference on Multimedia and Expo. ICME2000. Proceedings. Latest Advances in the Fast Changing World of Multimedia (Cat. No.00TH8532). Vol. 3. July 2000, 1641–1644 vol.3. doi: 10.1109/ICME.2000.871085 (cit. on p. 23).

[Bio14] Biocca, Frank: “Connected to My Avatar:” in: *Social Computing and Social Media.* Ed. by Meiselwitz, Gabriele. Cham: Springer International Publishing, 2014, pp. 421–429 (cit. on p. 13).

[Bof94] Boff, Kenneth R; Kaufman, Lloyd and Thomas, James P: Handbook of perception and human performance. Volume 2. Cognitive processes and performance. Tech. rep. DTIC Document, 1994 (cit. on p. 14).

[Bol19] Boletsis, Costas and Kongsvik, Stian: “Text input in virtual reality: A preliminary evaluation of the drum-like vr keyboard”. In: *Technologies* 7.2 (2019), p. 31. url: https://doi.org/10.3390/technologies7020031 (cit. on p. 13).

[Bon13] Bonnard, Quentin; Lemaignan, Séverin; Zufferey, Guillaume; Mazzei, Andrea; Cuendet, Sébastien; Li, Nan; Özgür, Ayberk and Dillenbourg, Pierre: Chilitags 2: Robust Fiducial Markers for Augmented Reality and Robotics. 2013. url: http://chili.epfl.ch/software (cit. on p. 86).

[Bon35] Bonferroni, Carlo E: “Il calcolo delle assicurazioni su gruppi di teste”. In: *Studi in onore del professore salvatore ortu carboni* (1935), pp. 13–60 (cit. on p. 296).

[Bor19] Born, F.; Sykownik, P. and Masuch, M.: “Co-Located vs. Remote Gameplay: The Role of Physical Co-Presence in Multiplayer Room-Scale VR”. In: *2019 IEEE Conference on Games (CoG)*. 2019, pp. 1–8. doi: 10.1109/CIG.2019.8848001 (cit. on pp. 22, 23, 112).

[Bot98] Botvinick, Matthew and Cohen, Jonathan: “Rubber hands ‘feel’touch that eyes see”. In: *Nature* 391.6669 (1998), pp. 756–756. doi: 10.1038/35784. url: https://doi.org/10.1038/35784 (cit. on p. 13).

[Bow01a] Bowman, D. A. and Wingrave, C. A.: “Design and evaluation of menu systems for immersive virtual environments”. In: *Proceedings IEEE Virtual Reality 2001*. 2001, pp. 149–156. doi: 10.1109/VR.2001.913781 (cit. on p. 18).

[Bow01b] Bowman, Doug A; Ly, Vinh Q and Campbell, Joshua M: “Pinch keyboard: Natural text input for immersive virtual environments”. In: (2001) (cit. on p. 13).

[Bow01c] Bowman, Doug A and Wingrave, Chadwick A: “Design and evaluation of menu systems for immersive virtual environments”. In: *Virtual Reality, 2001. Proceedings. IEEE.* IEEE. 2001, pp. 149–156 (cit. on p. 31).

[Bow01d] Bowman, Doug A.; Kruijff, Ernst; LaViola, Joseph J. and Poupyrev, Ivan: “An Introduction to 3-D User Interface Design”. In: *Presence: Teleoper. Virtual Environ.* 10.1 (Feb. 2001), pp. 96–108. doi: 10.1162/105474601750182342. url: http://dx.doi.org/10.1162/105474601750182342 (cit. on pp. 10, 13, 20, 35, 112).

[Bow02] Bowman, Doug A; Rhoton, Christopher J and Pinho, Marcio S: “Text input techniques for immersive virtual environments: An empirical comparison”. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 46. 26. SAGE Publications Sage CA: Los Angeles, CA. 2002, pp. 2154–2158 (cit. on pp. 13, 14, 39).

[Bow04] Bowman, Doug; Kruijff, Ernst; LaViola Jr, Joseph J and Poupyrev, Ivan: 3D User Interfaces: Theory and Practice. Addison-Wesley, 2004 (cit. on pp. 10, 19).

[Bow08] Bowman, Doug A; Coquillart, Sabine; Froehlich, Bernd; Hirose, Michitaka; Kitamura, Yoshifumi; Kiyokawa, Kiyoshi and Stuerzlinger, Wolfgang: “3D user interfaces: new directions and perspectives.” In: *IEEE computer graphics and applications* 28.6 (2008), pp. 20–36 (cit. on pp. 12, 14, 19).

[Bow97a] Bowman, Doug A and Hodges, Larry F: “An Evaluation of Techniques for Grabbing and Manipulating Remote Objects in Immersive Virtual Environments”. In: *Proceedings of the 1997 Symposium on Interactive 3D Graphics.* ACM. 1997, 35–ff (cit. on p. 132).

[Bow97b] Bowman, Doug A. and Hodges, Larry F.: “An Evaluation of Techniques for Grabbing and Manipulating Remote Objects in Immersive Virtual Environments”. In: *Proceedings of the 1997 Symposium on Interactive 3D Graphics.* I3D ’97. Providence, Rhode Island, USA: Association for Computing Machinery, 1997, 35–ff. DOI: 10.1145/253284.253301 (cit. on pp. 3, 68).

[Bow98] BOWMAN, D. A.; WINEMAN, J.; HODGES, L. F. and ALLISON, D.: "Designing animal habitats within an immersive VE". In: *IEEE Computer Graphics and Applications* 18.5 (1998), pp. 9–13. DOI: 10.1109/38.708555 (cit. on pp. 18, 51).

[Bow99] BOWMAN, Doug A and HODGES, Larry F: "Formalizing the design, evaluation, and application of interaction techniques for immersive virtual environments". In: *Journal of Visual Languages & Computing* 10.1 (1999), pp. 37–53 (cit. on p. 14).

[Bra15] BRANCATI, N.; CAGGIANESE, G.; FRUCCI, M.; GALLO, L. and NERONI, P.: "Robust fingertip detection in egocentric vision under varying illumination conditions". In: *2015 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*. 2015, pp. 1–6. DOI: 10.1109/ICMEW.2015.7169798 (cit. on p. 11).

[Bro02] BROWN, Michael Wayne; MCINTYRE, Joseph Herbert; MOORE, Victor S; PAOLINI, Michael A and WINTERS, Scott Lee: System and method for conference call line drop recovery. US Patent US7003286B2. Oct. 2002 (cit. on p. 149).

[Bro06] BROWN, L. D. and HUA, H.: "Magic Lenses for augmented virtual environments". In: *IEEE Computer Graphics and Applications* 26.4 (2006), pp. 64–73. DOI: 10.1109/MCG.2006.84 (cit. on p. 16).

[Bro99] BROOKS, Frederick P: "What's real about virtual reality?" In: *IEEE Computer graphics and applications* 19.6 (1999), pp. 16–27 (cit. on pp. 9, 10).

[Buc19] BUCK, L. E.; RIESER, J. J.; NARASIMHAM, G. and BODENHEIMER, B.: "Interpersonal Affordances and Social Dynamics in Collaborative Immersive Virtual Environments: Passing Together Through Apertures". In: *IEEE Transactions on Visualization and Computer Graphics* 25.5 (2019), pp. 2123–2133. DOI: 10.1109/TVCG.2019.2899232 (cit. on p. 25).

[Buc20] Buck, L. E.; Park, S. and Bodenheimer, B.: "Determining Peripersonal Space Boundaries and Their Plasticity in Relation to Object and Agent Characteristics in an Immersive Virtual Environment". In: *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 2020, pp. 332–342. doi: 10.1109/VR46266.2020.00053 (cit. on p. 30).

[Bue03] Bues, Matthias; Blach, Roland and Haselberger, Frank: "Sensing Surfaces: Bringing the Desktop into Virtual Environments". In: *Proceedings of the Workshop on Virtual Environments 2003*. EGVE '03. Zurich, Switzerland: Association for Computing Machinery, 2003, pp. 303–304. doi: 10.1145/769953.769989. url: https://doi.org/10.1145/769953.769989 (cit. on p. 21).

[Bur96] Burdea, Grigore C and Brooks, Frederick P: Force and touch feedback for virtual reality. Wiley New York, 1996 (cit. on p. 10).

[Büs19] Büschel, Wolfgang; Mitschick, Annett; Meyer, Thomas and Dachselt, Raimund: "Investigating Smartphone-Based Pan and Zoom in 3D Data Spaces in Augmented Reality". In: *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services*. MobileHCI '19. Taipei, Taiwan: Association for Computing Machinery, 2019. doi: 10.1145/3338286.3340113. url: https://doi.org/10.1145/3338286.3340113 (cit. on p. 15).

[Byr02] Byrne, Michael D: "Reading vertical text: Rotated vs. marquee". In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 46. 17. SAGE Publications Sage CA: Los Angeles, CA. 2002, pp. 1633–1635 (cit. on pp. 28, 130).

[Cad16] Cadena, A.; Carvajal, R.; Guamán, B.; Granda, R.; Peláez, E. and Chiluiza, K.: "Fingertip detection approach on depth image sequences for interactive projection system". In: *2016 IEEE Ecuador Technical Chapters Meeting (ETCM)*. 2016, pp. 1–6. doi: 10.1109/ETCM.2016.7750827 (cit. on p. 18).

[Cai18] Cai, Minghao; Masuko, Soh and Tanaka, Jiro: “Gesture-Based Mobile Communication System Providing Side-by-Side Shopping Feeling”. In: *Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion.* ACM. 2018 (cit. on p. 27).

[Cai19] Cai, Minghao and Tanaka, Jiro: “Go Together: Providing Nonverbal Awareness Cues to Enhance Co-Located Sensation in Remote Communication”. In: *Hum.-Centric Comput. Inf. Sci.* 9.1 (Dec. 2019). doi: 10.1186/s13673-019-0180-y. url: https://doi.org/10.1186/s13673-019-0180-y (cit. on p. 30).

[Cam20] Campbell, Abraham G.; Holz, Thomas; Cosgrove, Jonny; Harlick, Mike and O’Sullivan, Tadhg: “Uses of Virtual Reality for Communication in Financial Services: A Case Study on Comparing Different Telepresence Interfaces: Virtual Reality Compared to Video Conferencing”. In: *Advances in Information and Communication.* Ed. by Arai, Kohei and Bhatia, Rahul. Cham: Springer International Publishing, 2020, pp. 463–481 (cit. on p. 27).

[Car16] Carvalheiro, Cristiano; Nóbrega, Rui; Silva, Hugo da and Rodrigues, Rui: “User Redirection and Direct Haptics in Virtual Environments”. In: *Proceedings of the 24th ACM International Conference on Multimedia.* ACM. 2016, pp. 1146–1155 (cit. on p. 31).

[Cas16] Caserman, P.; Krabbe, P.; Wojtusch, J. and von Stryk, O.: “Real-time step detection using the integrated sensors of a head-mounted display”. In: *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC).* 2016, pp. 003510–003515. doi: 10.1109/SMC.2016.7844777 (cit. on p. 103).

[Cha12] Chan, Leith KY and Lau, Henry YK: “MagicPad: the projection based 3D user interface”. In: *International Journal on Interactive Design and Manufacturing (IJIDeM)* 6.2 (2012), pp. 75–81. url: https://doi.org/10.1007/s12008-012-0141-4 (cit. on p. 17).

[Che04] Chen, Jian; Bowman, Doug A.; Lucas, John F. and Wingrave, Chadwick A.: “Interfaces for Cloning in Immersive Virtual Environments”. In: *Proceedings of the Tenth Eurographics Conference on Virtual Environments*. EGVE’04. Grenoble, France: Eurographics Association, 2004, pp. 91–98 (cit. on p. 18).

[Che14a] Chen, Weiya; Clavel, Céline; Férey, Nicolas and Bourdot, Patrick: “Perceptual Conflicts in a Multi-Stereoscopic Immersive Virtual Environment: Case Study on Face-to-Face Interaction Through an Avatar”. In: *Presence: Teleoperators and Virtual Environments* 23.4 (2014), pp. 410–429 (cit. on p. 32).

[Che14b] Chen, Xiang’Anthony’; Grossman, Tovi and Fitzmaurice, George: “Swipeboard: a text entry technique for ultra-small interfaces that supports novice to expert transitions”. In: *Proceedings of the 27th annual ACM symposium on User interface software and technology*. ACM. 2014, pp. 615–620 (cit. on p. 13).

[Che16a] Chen, Ke-Yu; Patel, Shwetak N. and Keller, Sean: “Finexus: Tracking Precise Motions of Multiple Fingertips Using Magnetic Sensing”. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. CHI ’16. San Jose, California, USA: Association for Computing Machinery, 2016, pp. 1504–1514. doi: 10.1145/2858036.2858125. url: https://doi.org/10.1145/2858036.2858125 (cit. on p. 11).

[Che16b] Chethana, N. S.; Divyaprabha and Kurian, M. Z.: “Design and Implementation of Static Hand Gesture Recognition System for Device Control”. In: *Emerging Research in Computing, Information, Communication and Applications*. Ed. by Shetty, N. R.; Prasad, N Hamsavath and Nalini, N. Singapore: Springer Singapore, 2016, pp. 589–596 (cit. on p. 48).

[Che17] Cheng, Lung-Pan; Ofek, Eyal; Holz, Christian; Benko, Hrvoje and Wilson, Andrew D.: “Sparse Haptic Proxy: Touch Feedback in Virtual Environments Using a General Passive Prop”. In: *Proceedings of the 2017 CHI Conference on Human Factors* *in Computing Systems*. CHI '17. Denver, Colorado, USA: Association for Computing Machinery, 2017, pp. 3718–3728. DOI: 10.1145/3025453.3025753 (cit. on p. 31).

[Chi12] CHINELLO, F.; MALVEZZI, M.; PACCHIEROTTI, C. and PRATTICHIZZO, D.: "A three DoFs wearable tactile display for exploration and manipulation of virtual objects". In: *2012 IEEE Haptics Symposium (HAPTICS)*. 2012, pp. 71–76. DOI: 10.1109/HAPTIC.2012.6183772 (cit. on p. 16).

[Chl05] CHLOND, Martin J: "Box-Packing Puzzles". In: *INFORMS Transactions on Education* 5.3 (2005), pp. 70–72 (cit. on p. 138).

[Cho18] CHOI, Inrak; OFEK, Eyal; BENKO, Hrvoje; SINCLAIR, Mike and HOLZ, Christian: "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*. CHI '18. Montreal QC, Canada: Association for Computing Machinery, 2018, pp. 1–13. DOI: 10.1145/3173574.3174228. URL: https://doi.org/10.1145/3173574.3174228 (cit. on p. 16).

[Cig03] CIGER, Jan; GUTIERREZ, Mario; VEXO, Frederic and THALMANN, Daniel: "The magic wand". In: *Proceedings of the 19th spring conference on Computer graphics*. ACM. 2003, pp. 119–124 (cit. on p. 13).

[Cir09] CIRIO, Gabriel; MARCHAL, Maud; REGIA-CORTE, Tony and LÉCUYER, Anatole: "The Magic Barrier Tape: A Novel Metaphor for Infinite Navigation in Virtual Worlds with a Restricted Walking Workspace". In: *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology*. VRST '09. Kyoto, Japan: Association for Computing Machinery, 2009, pp. 155–162. DOI: 10.1145/1643928.1643965. URL: https://doi.org/10.1145/1643928.1643965 (cit. on p. 24).

[Cir12] CIRIO, G.; VANGORP, P.; CHAPOULIE, E.; MARCHAL, M.; LECUYER, A. and DRETTAKIS, G.: "Walking in a Cube: Novel Metaphors for Safely Navigating Large Virtual Environments in Restricted

Real Workspaces”. In: *IEEE Transactions on Visualization and Computer Graphics* 18.4 (2012), pp. 546–554. doi: 10.1109/TVCG.2012.60 (cit. on p. 24).

[Cla91] Clark, Herbert H and Brennan, Susan E: “Grounding in Communication”. In: *Perspectives on Socially Shared Cognition* 13.1991 (1991), pp. 127–149 (cit. on p. 29).

[Coh73] Cohen, Jacob: “Eta-squared and partial eta-squared in fixed factor ANOVA designs”. In: *Educational and psychological measurement* 33.1 (1973), pp. 107–112 (cit. on p. 298).

[Coh88] Cohen, Jacob: Statistical power analysis for the behavioral sciences. (2nd ed.) 1988 (cit. on p. 298).

[Con97] Coninx, K.; Van Reeth, F. and Flerackers, E.: “A hybrid 2D/3D user interface for immersive object modeling”. In: *Proceedings Computer Graphics International.* 1997, pp. 47–55. doi: 10.1109/CGI.1997.601270 (cit. on p. 20).

[Cor05] Correll, Kendall; Barendt, Nick and Branicky, Michael: “Design considerations for software only implementations of the IEEE 1588 precision time protocol”. In: *Conference on IEEE.* Vol. 1588. 2005, pp. 11–15 (cit. on p. 82).

[Cor13] Corsten, Christian; Avellino, Ignacio; Möllers, Max and Borchers, Jan: “Instant User Interfaces: Repurposing Everyday Objects as Input Devices”. In: *Proceedings of the 2013 ACM International Conference on Interactive Tabletops and Surfaces.* ITS ’13. St. Andrews, Scotland, United Kingdom: Association for Computing Machinery, 2013, pp. 71–80. doi: 10.1145/2512349.2512799. url: https://doi.org/10.1145/2512349.2512799 (cit. on p. 16).

[Cos20] Costes, A.; Danieau, F.; Argelaguet, F.; Guillotel, P. and Lécuyer, A.: “Towards Haptic Images: A Survey on Touchscreen-Based Surface Haptics”. In: *IEEE Transactions on Haptics* 13.3 (2020), pp. 530–541. doi: 10.1109/TOH.2020.2984754 (cit. on p. 16).

[Cze02] Czerwinski, Mary; Tan, Desney S and Robertson, George G: “Women take a wider view”. In: *Proceedings of the SIGCHI conference on Human factors in computing systems.* ACM. 2002, pp. 195–202 (cit. on p. 12).

[Dac07] Dachselt, Raimund and Hübner, Anett: “Three-dimensional menus: A survey and taxonomy”. In: *Computers & Graphics* 31.1 (2007), pp. 53–65. doi: https://doi.org/10.1016/j.cag.2006.09.006. url: http://www.sciencedirect.com/science/article/pii/S0097849306001853 (cit. on p. 20).

[Dai12] Dai, J. and Chung, R.: “Making any planar surface into a touch-sensitive display by a mere projector and camera”. In: *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops.* 2012, pp. 35–42. doi: 10.1109/CVPRW.2012.6239197 (cit. on p. 18).

[Dai14] Dai, J. and Chung, C. R.: “Touchscreen Everywhere: On Transferring a Normal Planar Surface to a Touch-Sensitive Display”. In: *IEEE Transactions on Cybernetics* 44.8 (2014), pp. 1383–1396. doi: 10.1109/TCYB.2013.2284512 (cit. on p. 18).

[Dan03] Dandekar, Kiran; Raju, Balasundar I. and Srinivasan, Mandayam A.: “3-D Finite-Element Models of Human and Monkey Fingertips to Investigate the Mechanics of Tactile Sense ”. In: *Journal of Biomechanical Engineering* 125.5 (Oct. 2003), pp. 682–691. doi: 10.1115/1.1613673 (cit. on p. 65).

[Dao18] Daoud, Hayyan Afeef; Md. Sabri, Aznul Qalid; Loo, Chu Kiong and Mansoor, Ali Mohammed: “SLAMM: Visual Monocular SLAM with Continuous Mapping Using Multiple Maps”. In: *PLoS ONE* 13.4 (Apr. 27, 2018). doi: 10.1371/journal.pone.0195878. pmid: 29702697 (cit. on pp. 11, 24).

[Dar05] Darken, Rudolph P. and Durost, Richard: “Mixed-Dimension Interaction in Virtual Environments”. In: *Proceedings of the ACM Symposium on Virtual Reality Software and Technology.* VRST ’05. Monterey, CA, USA: Association for Computing Machinery, 2005, pp. 38–45. doi: 10.1145/1101616.1101625. url: https://doi.org/10.1145/1101616.1101625 (cit. on pp. 19, 67, 170).

[Dem99] Dempski, Kelly L.: "Augmented Workspace: The World as Your Desktop". In: *Handheld and Ubiquitous Computing*. Ed. by Gellersen, Hans-W. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999, pp. 356–358 (cit. on p. 20).

[Dez12] Dezfuli, Niloofar; Khalilbeigi, Mohammadreza; Huber, Jochen; Müller, Florian and Mühlhäuser, Max: "PalmRC: Imaginary Palm-Based Remote Control for Eyes-Free Television Interaction". In: *Proceedings of the 10th European Conference on Interactive TV and Video*. EuroITV '12. Berlin, Germany: Association for Computing Machinery, 2012, pp. 27–34. doi: 10.1145/2325616.2325623. url: https://doi.org/10.1145/2325616.2325623 (cit. on p. 16).

[Di 03] Di Verdi, S.; Nurmi, D. and Hollerer, T.: "ARWin - a desktop augmented reality Window Manager". In: *The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003. Proceedings.* 2003, pp. 298–299. doi: 10.1109/ISMAR.2003.1240729 (cit. on p. 19).

[Dip11] Dippon, Andreas and Klinker, Gudrun: "KinectTouch: Accuracy Test for a Very Low-Cost 2.5D Multitouch Tracking System". In: *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*. ITS '11. Kobe, Japan: Association for Computing Machinery, 2011, pp. 49–52. doi: 10.1145/2076354.2076363. url: https://doi.org/10.1145/2076354.2076363 (cit. on p. 18).

[Dip17] Dipak, K.C. and Dawadi, Babu Ram: "Packet Loss Recovery and Control for VoIP". In: (2017), pp. 1804–1808 (cit. on p. 149).

[Dob15] Dobbelstein, David; Hock, Philipp and Rukzio, Enrico: "Belt: An Unobtrusive Touch Input Device for Head-Worn Displays". In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI '15. Seoul, Republic of Korea: Association for Computing Machinery, 2015, pp. 2135–2138.

DOI: 10.1145/2702123.2702450. URL: https://doi.org/10.1145/2702123.2702450 (cit. on p. 16).

[Dod10] DODDS, T; MOHLER, B and BÜLTHOFF, H: “A communication task in hmd virtual environments: Speaker and listener movement improves communication”. In: *23rd Annual Conference on Computer Animation and Social Agents (CASA 2010)*. 2010, pp. 1–4 (cit. on p. 3).

[Dod11] DODDS, Trevor J.; MOHLER, Betty J. and BÜLTHOFF, Heinrich H.: “Talk to the Virtual Hands: Self-Animated Avatars Improve Communication in Head-Mounted Display Virtual Environments”. In: *PLOS ONE* 6.10 (Oct. 2011), pp. 1–12. DOI: 10.1371/journal.pone.0025759. URL: https://doi.org/10.1371/journal.pone.0025759 (cit. on p. 32).

[Don19a] DONG, T.; SONG, Y.; SHEN, Y. and FAN, J.: “Simulation and Evaluation of Three-User Redirected Walking Algorithm in Shared Physical Spaces”. In: *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 2019, pp. 894–895. DOI: 10.1109/VR.2019.8798319 (cit. on p. 25).

[Don19b] DONG, Tianyang; SONG, Yifan; SHEN, Yuqi and FAN, Jing: “A Novel Method of Multi-user Redirected Walking for Large-Scale Virtual Environments”. In: *Advances in Computer Graphics*. Ed. by GAVRILOVA, Marina; CHANG, Jian; THALMANN, Nadia Magnenat; HITZER, Eckhard and ISHIKAWA, Hiroshi. Cham: Springer International Publishing, 2019, pp. 143–154 (cit. on p. 25).

[Don20] DONG, T.; CHEN, X.; SONG, Y.; YING, W. and FAN, J.: “Dynamic Artificial Potential Fields for Multi-User Redirected Walking”. In: *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 2020, pp. 146–154. DOI: 10.1109/VR46266.2020.00033 (cit. on p. 25).

[Dor16] DORTA, Tomás; KINAYOGLU, Gokce and HOFFMANN, Michael: “Hyve-3D and the 3D Cursor: Architectural co-design with freedom in Virtual Reality”. In: *International Journal of Architectural Computing* 14.2 (2016), pp. 87–102. DOI:

10.1177/1478077116638921. URL: https://doi.org/10.1177/1478077116638921 (cit. on p. 15).

[Dre20] Drey, Tobias; Gugenheimer, Jan; Karlbauer, Julian; Milo, Maximilian and Rukzio, Enrico: “VRSketchIn: Exploring the Design Space of Pen and Tablet Interaction for 3D Sketching in Virtual Reality”. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* CHI ’20. Honolulu, HI, USA: Association for Computing Machinery, 2020, pp. 1–14. DOI: 10.1145/3313831.3376628. URL: https://doi.org/10.1145/3313831.3376628 (cit. on p. 15).

[Enc99] Encarnaĉão, L. M.; Bimber, O.; Schmalstieg, D. and Chandler, S. D.: “A Translucent Sketchpad for the Virtual Table Exploring Motion-based Gesture Recognition”. In: *Computer Graphics Forum* 18.3 (1999), pp. 277–286. DOI: https://doi.org/10.1111/1467-8659.00348. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-8659.00348. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-8659.00348 (cit. on p. 17).

[Eng19] Englmeier, D.; Schönewald, I.; Butz, A. and Höllerer, T.: “Feel the Globe: Enhancing the Perception of Immersive Spherical Visualizations with Tangible Proxies”. In: *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR).* 2019, pp. 1693–1698. DOI: 10.1109/VR.2019.8797869 (cit. on pp. 16, 18).

[Eng68] Engelbart, Douglas C: “The mother of all demos”. In: *Fall Joint Computer Conference, San Francisco, Dec.* Vol. 9. 1968 (cit. on p. 1).

[Ens16] Ens, Barrett; Byagowi, Ahmad; Han, Teng; Hincapié-Ramos, Juan David and Irani, Pourang: “Combining Ring Input with Hand Tracking for Precise, Natural Interaction with Spatial Analytic Interfaces”. In: *Proceedings of the 2016 Symposium on Spatial User Interaction.* SUI ’16. Tokyo, Japan: Association for Computing Machinery, 2016, pp. 99–102. DOI: 10.1145/2983310.

2985757. URL: https://doi.org/10.1145/2983310.2985757 (cit. on p. 16).

[Eva99] Evans, Francine; Skiena, Steven and Varshney, Amitabh: “VType: Entering text in a virtual world”. In: *submitted to International Journal of Human-Computer Studies* (1999) (cit. on p. 14).

[Fal20] Faltaous, Sarah; Neuwirth, Joshua; Gruenefeld, Uwe and Schneegass, Stefan: “SaVR: Increasing Safety in Virtual Reality Environments via Electrical Muscle Stimulation”. In: *19th International Conference on Mobile and Ubiquitous Multimedia.* MUM 2020. Essen, Germany: Association for Computing Machinery, 2020, pp. 254–258. DOI: 10.1145/3428361.3428389. URL: https://doi.org/10.1145/3428361.3428389 (cit. on p. 24).

[Fec16] Fechteler, Philipp; Paier, Wolfgang; Hilsmann, Anna and Eisert, Peter: “Real-time avatar animation with dynamic face texturing”. In: *Image Processing (ICIP), 2016 IEEE International Conference on.* IEEE. 2016, pp. 355–359 (cit. on p. 12).

[Fei18] Feick, Martin; Mok, Terrance; Tang, Anthony; Oehlberg, Lora and Sharlin, Ehud: “Perspective on and Re-Orientation of Physical Proxies in Object-Focused Remote Collaboration”. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.* CHI ’18. Montreal QC, Canada: Association for Computing Machinery, 2018. DOI: 10.1145/3173574.3173855 (cit. on pp. 5, 30).

[Fei20] Feigl, T.; Gruner, L.; Mutschler, C. and Roth, D.: “Real-Time Gait Reconstruction For Virtual Reality Using a Single Sensor”. In: *2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct).* 2020, pp. 84–89. DOI: 10.1109/ISMAR-Adjunct51615.2020.00037 (cit. on p. 103).

[Fei93] Feiner, Steven; MacIntyre, Blair; Haupt, Marcus and Solomon, Eliot: “Windows on the World: 2D Windows for 3D Augmented Reality”. In: *Proceedings of the 6th Annual ACM Symposium on User Interface Software and Technology.* UIST

’93. Atlanta, Georgia, USA: Association for Computing Machinery, 1993, pp. 145–155. doi: 10.1145/168642.168657. url: https://doi.org/10.1145/168642.168657 (cit. on pp. 12, 21).

[Fer92] Ferraro, F Richard and Kellas, George: “Age-related changes in the effects of target orientation on word recognition”. In: *Journal of gerontology* 47.4 (1992), P279–P280 (cit. on pp. 28, 130).

[Fes50] Festinger, Leon; Schachter, Stanley and Back, Kurt: “Social Pressures in Informal Groups; A Study of Human Factors in Housing”. In: (1950) (cit. on p. 30).

[Feu17a] Feuchtner, Tiare and Müller, Jörg: “Extending the Body for Interaction with Reality”. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems.* ACM. 2017, pp. 5145–5157 (cit. on p. 31).

[Feu17b] Feuchtner, Tiare and Müller, Jörg: “Extending the Body for Interaction with Reality”. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems.* CHI ’17. Denver, Colorado, USA: Association for Computing Machinery, 2017, pp. 5145–5157. doi: 10.1145/3025453.3025689. url: https://doi.org/10.1145/3025453.3025689 (cit. on p. 31).

[Feu18] Feuchtner, Tiare and Müller, Jörg: “Ownershift: Facilitating Overhead Interaction in Virtual Reality with an Ownership-Preserving Hand Space Shift”. In: *The 31st Annual ACM Symposium on User Interface Software and Technology.* ACM. 2018, pp. 31–43 (cit. on p. 31).

[Fis92] Fisher, R. A.: “Statistical Methods for Research Workers”. In: *Breakthroughs in Statistics: Methodology and Distribution.* Ed. by Kotz, Samuel and Johnson, Norman L. New York, NY: Springer New York, 1992, pp. 66–70. doi: 10.1007/978-1-4612-4380-9_6. url: https://doi.org/10.1007/978-1-4612-4380-9_6 (cit. on p. 296).

[Fra99] FRASER, Mike; BENFORD, Steve; HINDMARSH, Jon and HEATH, Christian: “Supporting Awareness and Interaction Through Collaborative Virtual Interfaces”. In: *Proceedings of the 12th Annual ACM Symposium on User Interface Software and Technology.* ACM. 1999, pp. 27–36 (cit. on p. 27).

[Fré04] FRÉCON, Emmanuel and AVATARE-NÖU, Anneli: “Building Distributed Virtual Environments to Support Collaborative Work”. In: *1998 ACM Symposium on Virtual Reality Software and Technology (VRST’98), Taiwan.* 2004 (cit. on p. 3).

[Fré98] FRÉCON, Emmanuel and STENIUS, Mårten: “DIVE: A scaleable network architecture for distributed virtual environments”. In: *Distributed Systems Engineering* 5.3 (1998), p. 91 (cit. on p. 22).

[Fuj19] FUJINAWA, Eisuke; GOTO, Kenji; IRIE, Atsushi; WU, Songtao and XU, Kuanhong: “Occlusion-Aware Hand Posture Based Interaction on Tabletop Projector”. In: *The Adjunct Publication of the 32nd Annual ACM Symposium on User Interface Software and Technology.* UIST ’19. New Orleans, LA, USA: Association for Computing Machinery, 2019, pp. 113–115. DOI: 10.1145/3332167.3356890. URL: https://doi.org/10.1145/3332167.3356890 (cit. on pp. 17, 18).

[Fus00] FUSSELL, Susan R; KRAUT, Robert E and SIEGEL, Jane: “Coordination of Communication: Effects of Shared Visual Context on Collaborative Work”. In: *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work.* ACM. 2000, pp. 21–30 (cit. on pp. 27, 29, 30, 140).

[Fus03a] FUSSELL, Susan R; SETLOCK, Leslie D and PARKER, Elizabeth M: “Where Do Helpers Look?: Gaze Targets During Collaborative Physical Tasks”. In: *CHI’03 Extended Abstracts on Human Factors in Computing Systems.* ACM. 2003, pp. 768–769 (cit. on p. 138).

[Fus03b] FUSSELL, Susan R; SETLOCK, Leslie D; PARKER, Elizabeth M and YANG, Jie: “Assessing the Value of a Cursor Pointing Device for Remote Collaboration on Physical Tasks”. In: *CHI’03 Extended*

Abstracts on Human Factors in Computing Systems. ACM. 2003, pp. 788–789 (cit. on pp. 27, 140).

[Gal13] Galati, Alexia; Michael, Christina; Mello, Catherine; Greenauer, Nathan M and Avraamides, Marios N: “The conversational partner’s perspective affects spatial memory and descriptions”. In: *Journal of Memory and Language* 68.2 (2013), pp. 140–159 (cit. on p. 30).

[Gal15] Galati, Alexia and Avraamides, Marios N: “Social and representational cues jointly influence spatial perspective-taking”. In: *Cognitive science* 39.4 (2015), pp. 739–765 (cit. on p. 30).

[Gar01] Garau, Maia; Slater, Mel; Bee, Simon and Sasse, Martina Angela: “The Impact of Eye Gaze on Communication Using Humanoid Avatars”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM. 2001, pp. 309–316 (cit. on pp. 31, 140).

[Gar03] Garau, Maia; Slater, Mel; Vinayagamoorthy, Vinoba; Brogni, Andrea; Steed, Anthony and Sasse, M Angela: “The Impact of Avatar Realism and Eye Gaze Control on Perceived Quality of Communication in a Shared Immersive Virtual Environment”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM. 2003, pp. 529–536 (cit. on p. 32).

[Gar14] Garrido-Jurado, S.; Muñoz-Salinas, R.; Madrid-Cuevas, F.J. and Marín-Jiménez, M.J.: “Automatic generation and detection of highly reliable fiducial markers under occlusion”. In: *Pattern Recognition* 47.6 (2014), pp. 2280–2292. doi: 10.1016/j.patcog.2014.01.005 (cit. on p. 47).

[Gas19] Gasques, Danilo; Johnson, Janet G.; Sharkey, Tommy and Weibel, Nadir: “PintAR: Sketching Spatial Experiences in Augmented Reality”. In: *Companion Publication of the 2019 on Designing Interactive Systems Conference 2019 Companion.* DIS ’19 Companion. San Diego, CA, USA: Association for Computing

Machinery, 2019, pp. 17–20. DOI: 10.1145/3301019.3325158. URL: https://doi.org/10.1145/3301019.3325158 (cit. on p. 15).

[Gau14] Gauglitz, Steffen; Nuernberger, Benjamin; Turk, Matthew and Höllerer, Tobias: "World-Stabilized Annotations and Virtual Scene Navigation for Remote Collaboration". In: *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*. UIST '14. Honolulu, Hawaii, USA: Association for Computing Machinery, 2014, pp. 449–459. DOI: 10.1145/2642918.2647372 (cit. on p. 27).

[Gen13] Genest, Aaron M.; Gutwin, Carl; Tang, Anthony; Kalyn, Michael and Ivkovic, Zenja: "KinectArms: A Toolkit for Capturing and Displaying Arm Embodiments in Distributed Tabletop Groupware". In: *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*. CSCW '13. San Antonio, Texas, USA: Association for Computing Machinery, 2013, pp. 157–166. DOI: 10.1145/2441776.2441796. URL: https://doi.org/10.1145/2441776.2441796 (cit. on p. 27).

[Ger01] Gerhard, Michael; Moore, David J and Hobbs, Dave J: "Continuous presence in collaborative virtual environments: Towards a hybrid avatar-agent model for user representation". In: *International Workshop on Intelligent Virtual Agents*. Springer. 2001, pp. 137–155 (cit. on p. 104).

[Ger04] Gergle, Darren; Kraut, Robert E and Fussell, Susan R: "Language Efficiency and Visual Technology: Minimizing Collaborative Effort with Visual Information". In: *Journal of Language and Social Psychology* 23.4 (2004), pp. 491–517 (cit. on pp. 29, 138, 140).

[Gib79] Gibson, JJ: The ecological approach to visual perception. Boston, MA, US. 1979 (cit. on p. 12).

[Go08] Go, Kentaro; Konishi, Hayato and Matsuura, Yoshisuke: "Itone: a japanese text input method for a dual joystick game controller". In: *CHI'08 Extended Abstracts on Human Factors in*

Computing Systems. ACM. 2008, pp. 3141–3146 (cit. on pp. 13, 39, 40).

[Gol18] Gold, Omer and Sharir, Micha: “Dynamic Time Warping and Geometric Edit Distance: Breaking the Quadratic Barrier”. In: *ACM Trans. Algorithms* 14.4 (Aug. 2018). doi: 10.1145/3230734. url: https://doi.org/10.1145/3230734 (cit. on p. 82).

[Gom16] Gomes, Antonio; Rubens, Calvin; Braley, Sean and Vertegaal, Roel: “BitDrones: Towards Using 3D Nanocopter Displays as Interactive Self-Levitating Programmable Matter”. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.* CHI ’16. San Jose, California, USA: Association for Computing Machinery, 2016, pp. 770–780. doi: 10.1145/2858036.2858519. url: https://doi.org/10.1145/2858036.2858519 (cit. on p. 16).

[Góm16] Gómez Maureira, Marcello A. and Verbeek, Fons J.: “The Impact of Co-Located Play on Social Presence and Game Experience in a VR Game”. In: *DiGRA/FDG &#3916 - Proceedings of the First International Joint Conference of DiGRA and FDG.* Dundee, Scotland: Digital Games Research Association and Society for the Advancement of the Science of Digital Games, Aug. 2016. url: http://www.digra.org/wp-content/uploads/digital-library/paper_116.pdf (cit. on pp. 22, 23, 112).

[Gon09] González, Gabriel; Molina, José P; García, Arturo S; Martínez, Diego and González, Pascual: “Evaluation of text input techniques in immersive virtual environments”. In: *New Trends on Human–Computer Interaction.* Springer, 2009, pp. 109–118 (cit. on pp. 13, 14, 39, 40).

[Gon14] González-Franco, Mar; Peck, Tabitha C; Rodríguez-Fornells, Antoni and Slater, Mel: “A threat to a virtual hand elicits motor cortex activation”. In: *Experimental brain research* 232.3 (2014), pp. 875–887 (cit. on p. 13).

[Gon20] Gong, Jun; Gupta, Aakar and Benko, Hrvoje: “Acustico: Surface Tap Detection and Localization Using Wrist-Based Acoustic TDOA Sensing”. In: *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology.* UIST ’20. Virtual Event, USA: Association for Computing Machinery, 2020, pp. 406–419. doi: 10.1145/3379337.3415901. url: https://doi.org/10.1145/3379337.3415901 (cit. on p. 19).

[Gra17] Grandi, Jerônimo Gustavo; Debarba, Henrique Galvan; Nedel, Luciana and Maciel, Anderson: “Design and Evaluation of a Handheld-Based 3D User Interface for Collaborative Object Manipulation”. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems.* CHI ’17. Denver, Colorado, USA: Association for Computing Machinery, 2017, pp. 5881–5891. doi: 10.1145/3025453.3025935. url: https://doi.org/10.1145/3025453.3025935 (cit. on p. 15).

[Gra84] Graf, Peter and Levy, Betty Ann: “Reading and remembering: Conceptual and perceptual processing involved in reading rotated passages”. In: *Journal of Verbal Learning and Verbal Behavior* 23.3 (1984), pp. 405–424 (cit. on pp. 28, 130).

[Gre00] Greenhalgh, Chris; Purbrick, Jim and Snowdon, Dave: “Inside MASSIVE-3: Flexible support for data consistency and world structuring”. In: *Proceedings of the third international conference on Collaborative virtual environments.* ACM. 2000, pp. 119–127 (cit. on p. 21).

[Gre17a] Greenwald, Scott W; Corning, Wiley and Maes, Pattie: “Multi-User Framework for Collaboration and Co-Creation in Virtual Reality”. In: 12th International Conference on Computer Supported Collaborative Learning (CSCL). 2017 (cit. on p. 2).

[Gre17b] Greenwald, Scott W; Kulik, Alexander; Kunert, André; Beck, Stephan; Fröhlich, Bernd; Cobb, Sue; Parsons, Sarah; Newbutt, Nigel; Gouveia, Christine; Cook, Claire; Snyder, Anne; Payne, Scott; Holland, Jennifer; Buessing, Shawn; Fields, Gabriel; Corning, Wiley; Lee, Victoria; Xia, Lei and Maes, Pattie: “Technology and Applications for Collaborative Learning in Virtual Reality”. In: Philadelphia, PA: International Society of the Learning Sciences., 2017 (cit. on pp. 4, 21, 27).

[Gre91] Green, Mark and Jacob, Robert: “SIGGRAPH’90 Workshop report: software architectures and metaphors for non-WIMP user interfaces”. In: *ACM SIGGRAPH Computer Graphics* 25.3 (1991), pp. 229–235 (cit. on p. 9).

[Gro15] Grossman, Tovi; Chen, Xiang Anthony and Fitzmaurice, George: “Typing on glasses: Adapting text entry to smart eyewear”. In: *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services.* ACM. 2015, pp. 144–152 (cit. on pp. 13, 39).

[Gro18] Groeger, Daniel and Steimle, Jürgen: “ObjectSkin: Augmenting Everyday Objects with Hydroprinted Touch Sensors and Displays”. In: *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1.4 (Jan. 2018). doi: 10.1145/3161165. url: https://doi.org/10.1145/3161165 (cit. on p. 18).

[Gru15] Grubert, Jens; Pahud, Michel; Grasset, Raphael; Schmalstieg, Dieter and Seichter, Hartmut: “The utility of Magic Lens interfaces on handheld devices for touristic map navigation”. In: *Pervasive and Mobile Computing* 18 (2015), pp. 88–103. doi: https://doi.org/10.1016/j.pmcj.2014.08.005. url: http://www.sciencedirect.com/science/article/pii/S1574119214001370 (cit. on p. 16).

[Gru18a] Grubert, J.; Witzani, L.; Ofek, E.; Pahud, M.; Kranz, M. and Kristensson, P. O.: “Effects of Hand Representations for Typing in Virtual Reality”. In: *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR).* 2018, pp. 151–158. doi: 10.1109/VR.2018.8446250 (cit. on pp. 17, 39, 40, 45, 46, 66, 76).

[Gru18b] Grubert, J.; Witzani, L.; Ofek, E.; Pahud, M.; Kranz, M. and Kristensson, P. O.: “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). 2018, pp. 159–166. DOI: 10.1109/VR.2018.8446059 (cit. on p. 17).

[Gug16] GUGENHEIMER, Jan; DOBBELSTEIN, David; WINKLER, Christian; HAAS, Gabriel and RUKZIO, Enrico: "FaceTouch: Enabling Touch Interaction in Display Fixed UIs for Mobile Virtual Reality". In: *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. UIST '16. Tokyo, Japan: Association for Computing Machinery, 2016, pp. 49–60. DOI: 10.1145/2984511.2984576. URL: https://doi.org/10.1145/2984511.2984576 (cit. on p. 15).

[Gui87] GUIARD, Yves: "Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model". In: *Journal of motor behavior* 19.4 (1987), pp. 486–517 (cit. on p. 15).

[Ha14] HA, T.; FEINER, S. and WOO, W.: "WeARHand: Head-worn, RGB-D camera-based, bare-hand user interface with visually enhanced depth perception". In: *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 2014, pp. 219–228. DOI: 10.1109/ISMAR.2014.6948431 (cit. on p. 11).

[Haa06] HAAN, Gerwin de; GRIFFITH, Eric J; KOUTEK, Michal and POST, Frits H: "Hybrid Interfaces in VEs: Intent and Interaction". In: *EGVE*. 2006, pp. 109–118 (cit. on pp. 10, 19, 132).

[Hal00] HALLER, Cynthia R; GALLAGHER, Victoria J; WELDON, Tracey L and FELDER, Richard M: "Dynamics of Peer Education in Cooperative Learning Workgroups". In: *Journal of Engineering Education* 89.3 (2000), pp. 285–293 (cit. on p. 31).

[Han07] HANNA, Joy E and BRENNAN, Susan E: "Speakers' Eye Gaze Disambiguates Referring Expressions Early During Face-to-Face Conversation". In: *Journal of Memory and Language* 57.4 (2007), pp. 596–615 (cit. on p. 31).

[Han18] Han, Dustin T; Suhail, Mohamed and Ragan, Eric D: "Evaluating Remapped Physical Reach for Hand Interactions with Passive Haptics in Virtual Reality". In: *IEEE Transactions on Visualization and Computer Graphics* 24.4 (2018), pp. 1467–1476 (cit. on p. 31).

[Har04] Harms, Chad and Biocca, Frank: "Internal Consistency and Reliability of the Networked Minds Measure of Social Presence". In: *M. Alcaniz & B. Rey (Eds.), Seventh Annual International Workshop: Presence* (2004) (cit. on p. 293).

[Har06] Hart, Sandra G: "NASA-Task Load Index (NASA-TLX); 20 Years Later". In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 50. 9. Sage Publications. 2006, pp. 904–908 (cit. on p. 293).

[Har11] Harrison, Chris; Benko, Hrvoje and Wilson, Andrew D.: "OmniTouch: Wearable Multitouch Interaction Everywhere". In: *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*. UIST '11. Santa Barbara, California, USA: Association for Computing Machinery, 2011, pp. 441–450. doi: 10.1145/2047196.2047255. url: https://doi.org/10.1145/2047196.2047255 (cit. on pp. 16, 17, 55).

[Har12] Hardy, John and Alexander, Jason: "Toolkit Support for Interactive Projected Displays". In: *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia*. MUM '12. Ulm, Germany: Association for Computing Machinery, 2012. doi: 10.1145/2406367.2406419. url: https://doi.org/10.1145/2406367.2406419 (cit. on pp. 17, 18).

[Har17] Harley, Daniel; Tarun, Aneesh P.; Germinario, Daniel and Mazalek, Ali: "Tangible VR: Diegetic Tangible Objects for Virtual Reality Narratives". In: *Proceedings of the 2017 Conference on Designing Interactive Systems*. DIS '17. Edinburgh, United Kingdom: Association for Computing Machinery, 2017, pp. 1253–1263. doi: 10.1145/3064663.3064680. url: https://doi.org/10.1145/3064663.3064680 (cit. on p. 17).

[Haz05] Hazas, Mike; Kray, Christian; Gellersen, Hans; Agbota, Henoc; Kortuem, Gerd and Krohn, Albert: “A Relative Positioning System for Co-Located Mobile Devices”. In: *Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services.* MobiSys ’05. Seattle, Washington: Association for Computing Machinery, 2005, pp. 177–190. doi: 10.1145/1067170.1067190. url: https://doi.org/10.1145/1067170.1067190 (cit. on p. 24).

[Hei00] Heim, Michael: Virtual realism. Oxford University Press, 2000 (cit. on p. 9).

[Hei57] Heilig, Morton L: Stereoscopic-television apparatus for individual use. US Patent US2955156A. May 1957 (cit. on p. 1).

[Hel06] Heldal, Ilona; Spante, Maria and Connell, Mike: “Are two heads better than one?: object-focused work in physical and in virtual environments”. In: *Proceedings of the ACM symposium on Virtual reality software and technology.* ACM. 2006, pp. 287–296 (cit. on pp. 3, 22).

[Hel14] Helbig, Carolin; Bauer, Hans-Stefan; Rink, Karsten; Wulfmeyer, Volker; Frank, Michael and Kolditz, Olaf: “Concept and workflow for 3D visualization of atmospheric data in a virtual reality environment for analytical approaches”. In: *Environmental Earth Sciences* 72.10 (2014), p. 3767 (cit. on p. 2).

[Hep19] Hepperle, Daniel; Weiß, Yannick; Siess, Andreas and Wölfel, Matthias: “2D, 3D or speech? A case study on which user interface is preferable for what kind of object interaction in immersive virtual reality”. In: *Computers & Graphics* 82 (2019), pp. 321–331. doi: https://doi.org/10.1016/j.cag.2019.06.003. url: http://www.sciencedirect.com/science/article/pii/S0097849319300974 (cit. on p. 19).

[Her94] HERNDON, Kenneth P and MEYER, Tom: “3D widgets for exploratory scientific visualization”. In: *Proceedings of the 7th annual ACM symposium on User interface software and technology.* ACM. 1994, pp. 69–70 (cit. on p. 12).

[Hes11] HESSELMANN, Tobias; HEUTEN, Wilko and BOLL, Susanne: “Tap2Count: numerical input for interactive tabletops”. In: *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces.* ACM. 2011, pp. 256–257 (cit. on p. 14).

[Hig15] HIGUCHI, Keita; CHEN, Yinpeng; CHOU, Philip A; ZHANG, Zhengyou and LIU, Zicheng: “ImmerseBoard: Immersive telepresence experience using a digital whiteboard”. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems.* ACM. 2015, pp. 2383–2392 (cit. on p. 26).

[Hil16] HILD, Jutta; PETERSEN, Patrick and BEYERER, Jürgen: “Moving Target Acquisition by Gaze Pointing and Button Press Using Hand or Foot”. In: *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research and Applications.* ETRA ’16. Charleston, South Carolina: Association for Computing Machinery, 2016, pp. 257–260. DOI: 10.1145/2857491.2857535. URL: https://doi.org/10.1145/2857491.2857535 (cit. on p. 14).

[Hin18a] HINDERKS, Andreas; SCHREPP, Martin and THOMASCHEWSKI, Jörg: “A Benchmark for the Short Version of the User Experience Questionnaire”. In: Sept. 2018. DOI: 10.5220/0007188303730377 (cit. on p. 63).

[Hin18b] HINDERKS, Andreas; SCHREPP, Martin and THOMASCHEWSKI, Jörg: “A Benchmark for the Short Version of the User Experience Questionnaire”. In: *WEBIST.* 2018, pp. 373–377 (cit. on p. 146).

[Hin94] HINCKLEY, Ken; PAUSCH, Randy; GOBLE, John C and KASSELL, Neal F: “Passive real-world interface props for neurosurgical visualization”. In: *Proceedings of the SIGCHI conference on Human factors in computing systems.* ACM. 1994, pp. 452–458 (cit. on p. 15).

[Hin98] Hindmarsh, Jon; Fraser, Mike; Heath, Christian; Benford, Steve and Greenhalgh, Chris: "Fragmented Interaction: Establishing Mutual Orientation in Virtual Environments". In: *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work*. CSCW '98. Seattle, Washington, USA: Association for Computing Machinery, 1998, pp. 217–226. doi: 10.1145/289444.289496. url: https://doi.org/10.1145/289444.289496 (cit. on p. 32).

[Hoa16] Hoang, Thuong N; Reinoso, Martin; Vetere, Frank and Tanin, Egemen: "Onebody: Remote Posture Guidance System Using First Person View in Virtual Environment". In: *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*. ACM. 2016, p. 25 (cit. on p. 27).

[Hol14] Holman, David; Fellion, Nicholas and Vertegaal, Roel: "Sensing Touch Using Resistive Graphs". In: *Proceedings of the 2014 Conference on Designing Interactive Systems*. DIS '14. Vancouver, BC, Canada: Association for Computing Machinery, 2014, pp. 195–198. doi: 10.1145/2598510.2598552. url: https://doi.org/10.1145/2598510.2598552 (cit. on pp. 16, 18).

[Hop18a] Hoppe, Adrian H.; Otto, Leonard; van de Camp, Florian; Stiefelhagen, Rainer and Unmüßig, Gabriel: "qVRty: Virtual Keyboard with a Haptic, Real-World Representation". In: *HCI International 2018 – Posters' Extended Abstracts*. Cham: Springer International Publishing, 2018, pp. 266–272. doi: 10.1007/978-3-319-92279-9_36 (cit. on pp. 8, 10, 17, 30, 35).

[Hop18b] Hoppe, Adrian H.; Reeb, Roland; van de Camp, Florian and Stiefelhagen, Rainer: "Capability for Collision Avoidance of Different User Avatars in Virtual Reality". In: *HCI International 2018 – Posters' Extended Abstracts*. Cham: Springer International Publishing, 2018, pp. 273–279. doi: 10.1007/978-3-319-92279-9_37 (cit. on pp. 8, 10, 23, 79).

[Hop18c] Hoppe, Adrian H.; Reeb, Roland; van de Camp, Florian and Stiefelhagen, Rainer: "Interaction of Distant and Local Users in a Collaborative Virtual Environment". In: *Virtual, Augmented and Mixed Reality: Interaction, Navigation, Visualization, Embodiment, and Simulation.* Cham: Springer International Publishing, 2018, pp. 328–337. DOI: 10.1007/978-3-319-91581-4_24 (cit. on pp. 8, 10, 79).

[Hop18d] Hoppe, Adrian H.; van de Camp, Florian and Stiefelhagen, Rainer: "Personal Perspective: Using Modified World Views to Overcome Real-Life Limitations in Virtual Reality". In: *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR).* 2018, pp. 577–578. DOI: 10.1109/VR.2018.8446311 (cit. on pp. 8, 10, 116).

[Hop18e] Hoppe, Matthias; Knierim, Pascal; Kosch, Thomas; Funk, Markus; Futami, Lauren; Schneegass, Stefan; Henze, Niels; Schmidt, Albrecht and Machulla, Tonja: "VRHapticDrones: Providing Haptics in Virtual Reality through Quadcopters". In: *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia.* MUM 2018. Cairo, Egypt: Association for Computing Machinery, 2018, pp. 7–18. DOI: 10.1145/3282894.3282898. URL: https://doi.org/10.1145/3282894.3282898 (cit. on p. 16).

[Hop19] Hoppe, Adrian H.; Marek, Felix; van de Camp, Florian and Stiefelhagen, Rainer: "VirtualTablet: Extending Movable Surfaces with Touch Interaction". In: *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR).* Mar. 2019, pp. 980–981. DOI: 10.1109/VR.2019.8797993 (cit. on pp. 8, 10, 35).

[Hop20a] Hoppe, Adrian H.; Kaucher, Leon; van de Camp, Florian and Stiefelhagen, Rainer: "Calibration of Diverse Tracking Systems to Enable Local Collaborative Mixed Reality Applications". In: *Virtual, Augmented and Mixed Reality. Design and Interaction.* Cham: Springer International Publishing, 2020, pp. 63–80. DOI: 10.1007/978-3-030-49695-1_5 (cit. on pp. 8, 10, 79).

[Hop20b] Hoppe, Adrian H.; Marek, Felix; van de Camp, Florian and Stiefelhagen, Rainer: "Extending Movable Surfaces with

Touch Interaction Using the VirtualTablet : An Extended View". In: *Advances in Science, Technology and Engineering Systems Journal* 5.2 (2020), pp. 328–337. DOI: 10.25046/aj050243 (cit. on pp. 8, 10, 35).

[Hop20c] HOPPE, Adrian H.; VAN DE CAMP, Florian and STIEFELHAGEN, Rainer: "Enabling Interaction with Arbitrary 2D Applications in Virtual Environments". In: *HCI International 2020 - Posters.* Cham: Springer International Publishing, 2020, pp. 30–36. DOI: 10.1007/978-3-030-50729-9_4 (cit. on pp. 8, 10, 35).

[Hop21] HOPPE, Adrian H.; VAN DE CAMP, Florian and STIEFELHAGEN, Rainer: "ShiSha: Enabling Shared Perspective With Face-to-Face Collaboration Using Redirected Avatars in Virtual Reality". In: *Proc. ACM Hum.-Comput. Interact.* 4.CSCW 3 (Jan. 2021). DOI: 10.1145/3432950 (cit. on pp. 8, 10, 116).

[Hos12] HOSTE, Lode; DUMAS, Bruno and SIGNER, Beat: "SpeeG: a multimodal speech-and gesture-based text input solution". In: *Proceedings of the International working conference on advanced visual interfaces.* ACM. 2012, pp. 156–163 (cit. on pp. 14, 39).

[Hsu05] HSU, Henry and LACHENBRUCH, Peter A.: "Paired t Test". In: *Encyclopedia of Biostatistics.* American Cancer Society, 2005. DOI: https://doi.org/10.1002/0470011815.b2a15112. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/0470011815.b2a15112. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/0470011815.b2a15112 (cit. on p. 296).

[Hu14] HU, J.; LI, G.; XIE, X.; LV, Z. and WANG, Z.: "Bare-fingers Touch Detection by the Button's Distortion in a Projector–Camera System". In: *IEEE Transactions on Circuits and Systems for Video Technology* 24.4 (2014), pp. 566–575. DOI: 10.1109/TCSVT.2013.2280088 (cit. on p. 18).

[Hua13a] HUANG, Weidong and ALEM, Leila: "HandsinAir: A Wearable System for Remote Collaboration on Physical Tasks". In: *Proceedings of the 2013 Conference on Computer Supported Cooperative Work Companion.* ACM. 2013, pp. 153–156 (cit. on p. 138).

[Hua13b] Huang, Weidong; Alem, Leila and Tecchia, Franco: "HandsIn3D: Augmenting the Shared 3D Visual Space With Unmediated Hand Gestures". In: *SIGGRAPH Asia 2013 Emerging Technologies*. ACM. 2013, p. 10 (cit. on p. 28).

[Hua13c] Huang, Weidong; Alem, Leila and Tecchia, Franco: "HandsIn3D: Supporting Remote Guidance with Immersive Virtual Environments". In: *IFIP Conference on Human-Computer Interaction*. Springer. 2013, pp. 70–77 (cit. on p. 28).

[Hua18a] Huang, Shaoyan; Bai, Huidong; Mandalika, V B H and Lindeman, Robert W.: "Improving Virtual Reality Safety Precautions with Depth Sensing". In: *Proceedings of the 30th Australian Conference on Computer-Human Interaction*. OzCHI '18. Melbourne, Australia: Association for Computing Machinery, 2018, pp. 528–531. doi: 10.1145/3292147.3292241. url: https://doi.org/10.1145/3292147.3292241 (cit. on p. 24).

[Hua18b] Huang, Weidong; Alem, Leila; Tecchia, Franco and Duh, Henry Been-Lirn: "Augmented 3D Hands: A Gesture-Based Mixed Reality System for Distributed Collaboration". In: *Journal on Multimodal User Interfaces* 12.2 (2018), pp. 77–89 (cit. on p. 27).

[Huo17] Huo, Ke; Vinayak and Ramani, Karthik: "Window-Shaping: 3D Design Ideation by Creating on, Borrowing from, and Looking at the Physical World". In: *Proceedings of the Eleventh International Conference on Tangible, Embedded, and Embodied Interaction*. TEI '17. Yokohama, Japan: Association for Computing Machinery, 2017, pp. 37–45. doi: 10.1145/3024969.3024995. url: https://doi.org/10.1145/3024969.3024995 (cit. on p. 15).

[Huo18] Huo, Ke; Wang, Tianyi; Paredes, Luis; Villanueva, Ana M.; Cao, Yuanzhi and Ramani, Karthik: "SynchronizAR: Instant Synchronization for Spontaneous and Spatial Collaborations in Augmented Reality". In: *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*. UIST '18.

Berlin, Germany: ACM, 2018, pp. 19–30. DOI: 10.1145/3242587.3242595 (cit. on p. 23).

[Ish91] ISHII, Hiroshi and MIYAKE, Naomi: "Toward an open shared workspace: computer and video fusion approach of TeamWorkStation". In: *Communications of the ACM* 34.12 (1991), pp. 37–50 (cit. on p. 26).

[Ish92] ISHII, Hiroshi and KOBAYASHI, Minoru: "ClearBoard: a seamless medium for shared drawing and conversation with eye contact". In: *Proceedings of the SIGCHI conference on Human factors in computing systems.* ACM. 1992, pp. 525–532 (cit. on p. 26).

[Iso04] ISOKOSKI, Poika and RAISAMO, Roope: "Quikwriting as a multi-device text entry method". In: *Proceedings of the third Nordic conference on Human-computer interaction.* ACM. 2004, pp. 105–108 (cit. on pp. 13, 39, 40).

[Iza11] IZADI, Shahram; KIM, David; HILLIGES, Otmar; MOLYNEAUX, David; NEWCOMBE, Richard; KOHLI, Pushmeet; SHOTTON, Jamie; HODGES, Steve; FREEMAN, Dustin; DAVISON, Andrew and FITZGIBBON, Andrew: "KinectFusion: Real-Time 3D Reconstruction and Interaction Using a Moving Depth Camera". In: *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology.* UIST '11. Santa Barbara, California, USA: Association for Computing Machinery, 2011, pp. 559–568. DOI: 10.1145/2047196.2047270. URL: https://doi.org/10.1145/2047196.2047270 (cit. on p. 18).

[Jac92] JACOBY, Richard H. and ELLIS, Stephen R.: "Using virtual menus in a virtual environment". In: *Visual Data Interpretation.* Ed. by ALEXANDER, Joanna R. Vol. 1668. International Society for Optics and Photonics. SPIE, 1992, pp. 39–48. DOI: 10.1117/12.59654. URL: https://doi.org/10.1117/12.59654 (cit. on p. 19).

[Jai19] JAIN, Abhinav and DELLAERT, Frank: Fast 3D Pose Refinement with RGB Images. Nov. 17, 2019 (cit. on p. 91).

[Joh65] Johnson, E. A.: “Touch display–a novel input/output device for computers”. In: *Electronics Letters* 1.8 (1965), pp. 219–220. doi: 10.1049/el:19650200 (cit. on p. 1).

[Jos19] Joshi, Nikhita and Vogel, Daniel: “An Evaluation of Touch Input at the Edge of a Table”. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.* CHI ’19. Glasgow, Scotland Uk: Association for Computing Machinery, 2019, pp. 1–12. doi: 10.1145/3290605.3300476. url: https://doi.org/10.1145/3290605.3300476 (cit. on p. 17).

[Kak08] Kakehi, Yasuaki and Naemura, Takeshi: “UlteriorScape: Interactive optical superimposition on a view-dependent tabletop display”. In: *Horizontal Interactive Human Computer Systems, 2008. TABLETOP 2008. 3rd IEEE International Workshop on.* IEEE. 2008, pp. 189–192 (cit. on p. 29).

[Kal02] Kalkusch, M.; Lidy, T.; Knapp, N.; Reitmayr, G.; Kaufmann, H. and Schmalstieg, D.: “Structured Visual Markers for Indoor Pathfinding”. In: *The First IEEE International Workshop Agumented Reality Toolkit,* First IEEE International Augmented Reality Toolkit Workshop. Darmstadt, Germany: IEEE, 2002, p. 8. doi: 10.1109/ART.2002.1107018 (cit. on p. 23).

[Kal14] Kalckert, Andreas and Ehrsson, H. Henrik: “The spatial distance rule in the moving and classical rubber hand illusions”. In: *Consciousness and Cognition* 30 (2014), pp. 118–132. doi: https://doi.org/10.1016/j.concog.2014.08.022. url: http://www.sciencedirect.com/science/article/pii/S1053810014001603 (cit. on p. 31).

[Kam09] Kamuro, S.; Minamizawa, K.; Kawakami, N. and Tachi, S.: “Ungrounded kinesthetic pen for haptic interaction with virtual environments”. In: *RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication.* 2009, pp. 436–441. doi: 10.1109/ROMAN.2009.5326217 (cit. on p. 16).

[Kan19] Kang, HyeongYeop; Lee, Geonsun and Han, JunaHyun: “Visual Manipulation for Underwater Drag Force Perception in Immersive Virtual Environments”. In: *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR).* IEEE. 2019, pp. 38–46 (cit. on p. 31).

[Kar12] Karnik, Abhijit; Martinez Plasencia, Diego; Mayol-Cuevas, Walterio and Subramanian, Sriram: “PiVOT: personalized view-overlays for tabletops”. In: *Proceedings of the 25th annual ACM symposium on User interface software and technology.* ACM. 2012, pp. 271–280 (cit. on p. 29).

[Kas14] Kasahara, Shunichi and Rekimoto, Jun: “JackIn: Integrating First-Person View with Out-of-Body Vision Generation for Human-Human Augmentation”. In: *Proceedings of the 5th Augmented Human International Conference.* ACM. 2014, p. 46 (cit. on pp. 28, 29).

[Kas16] Kasahara, Shunichi; Nagai, Shohei and Rekimoto, Jun: “JackIn Head: Immersive Visual Telepresence System with Omnidirectional Wearable Camera”. In: *IEEE Transactions on Visualization and Computer Graphics* 23.3 (2016), pp. 1222–1234 (cit. on p. 28).

[Kat15] Katzakis, Nicholas; Teather, Robert J; Kiyokawa, Kiyoshi and Takemura, Haruo: “INSPECT: extending plane-casting for 6-DOF control”. In: *Human-centric Computing and Information Sciences* 5.1 (2015), pp. 1–22. url: https://doi.org/10.1186/s13673-015-0037-y (cit. on p. 15).

[Kec03] Keckeisen, M.; Stoev, S. L.; Feurer, M. and Strasser, W.: “Interactive cloth simulation in virtual environments”. In: *IEEE Virtual Reality, 2003. Proceedings.* 2003, pp. 71–78. doi: 10.1109/VR.2003.1191123 (cit. on p. 19).

[Kem18] Kemmoku, Yusuke and Komuro, Takashi: “A Large Effective Touchscreen Using a Head-Mounted Projector”. In: *Information* 9.9 (2018), p. 235. url: https://doi.org/10.3390/info9090235 (cit. on p. 17).

[Key13] Keyton, Joann; Caputo, Jennifer Marie; Ford, Emily Anne; Fu, Rong; Leibowitz, Samantha A; Liu, Tingting; Polasik, Sarah S; Ghosh, Paromita and Wu, Chaofan: "Investigating Verbal Workplace Communication Behaviors". In: *The Journal of Business Communication (1973)* 50.2 (2013), pp. 152–169 (cit. on p. 31).

[Kil12a] Kilteni, Konstantina; Groten, Raphaela and Slater, Mel: "The Sense of Embodiment in Virtual Reality". In: *Presence: Teleoperators and Virtual Environments* 21.4 (2012), pp. 373–387. doi: 10.1162/PRES_a_00124. url: https://doi.org/10.1162/PRES_a_00124 (cit. on p. 12).

[Kil12b] Kilteni, Konstantina; Normand, Jean-Marie; Sanchez-Vives, Maria V. and Slater, Mel: "Extending Body Space in Immersive Virtual Reality: A Very Long Arm Illusion". In: *PLOS ONE* 7.7 (July 2012), pp. 1–15. doi: 10.1371/journal.pone.0040867. url: https://doi.org/10.1371/journal.pone.0040867 (cit. on p. 31).

[Kil15] Kilteni, Konstantina; Maselli, Antonella; Kording, Konrad P. and Slater, Mel: "Over my fake body: body ownership illusions for studying the multisensory basis of own-body perception". In: *Frontiers in Human Neuroscience* 9 (2015), p. 141. doi: 10.3389/fnhum.2015.00141. url: https://www.frontiersin.org/article/10.3389/fnhum.2015.00141 (cit. on p. 13).

[Kim04] Kim, Sooyoung and Kim, Gerard Jounghyun: "Using keyboards with head mounted displays". In: *Proceedings of the 2004 ACM SIGGRAPH international conference on Virtual Reality continuum and its applications in industry.* ACM. 2004, pp. 336–343 (cit. on p. 13).

[Kim09] Kimani, Stephen: "WIMP Interfaces". In: *Encyclopedia of Database Systems.* Ed. by LIU, LING and ÖZSU, M. TAMER. Boston, MA: Springer US, 2009, pp. 3529–3533. doi: 10.1007/978-0-387-39940-9_467. url: https://doi.org/10.1007/978-0-387-39940-9_467 (cit. on p. 1).

[Kim13] Kim, Hyejin and Lee, Sung-Hee: "Reconstructing whole-body motions with wrist trajectories". In: *Graphical Models* 75.6 (2013), pp. 328–345. doi: https://doi.org/10.1016/j.gmod.2013.08.002. url: http://www.sciencedirect.com/science/article/pii/S152407031300026X (cit. on p. 103).

[Kim16] Kim, Hwan; Kim, Minhwan and Lee, Woohun: "HapThimble: A Wearable Haptic Device towards Usable Virtual Touch Screen". In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.* CHI '16. San Jose, California, USA: Association for Computing Machinery, 2016, pp. 3694–3705. doi: 10.1145/2858036.2858196. url: https://doi.org/10.1145/2858036.2858196 (cit. on p. 16).

[Kim17a] Kim, Mingyu; Jeon, Changyu and Kim, Jinmo: "A study on immersion and presence of a portable hand haptic system for immersive virtual reality". In: *Sensors* 17.5 (2017), p. 1141 (cit. on p. 16).

[Kim17b] Kim, Youngwon R and Kim, Gerard J: "HoVR-Type: Smartphone as a typing interface in VR using hovering". In: *Consumer Electronics (ICCE), 2017 IEEE International Conference on.* IEEE. 2017, pp. 200–203 (cit. on pp. 13, 39).

[Kim19a] Kim, Huhn; Yi, Seungyoun and Yoon, So-Yeon: "Exploring touch feedback display of virtual keyboards for reduced eye movements". In: *Displays* 56 (2019), pp. 38–48. doi: https://doi.org/10.1016/j.displa.2018.11.004. url: http://www.sciencedirect.com/science/article/pii/S0141938218301033 (cit. on p. 47).

[Kim19b] Kim, Ue-Hwan; Yoo, Sahng-Min and Kim, Jong-Hwan: "I-Keyboard: Fully Imaginary Keyboard on Touch Devices Empowered by Deep Neural Decoder". In: *IEEE transactions on cybernetics* (Nov. 2019). doi: 10.1109/tcyb.2019.2952391. url: https://doi.org/10.1109/TCYB.2019.2952391 (cit. on p. 17).

[Kir06] Kirk, David and Stanton Fraser, Danae: "Comparing Remote Gesture Technologies for Supporting Collaborative Physical

Tasks”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM. 2006, pp. 1191–1200 (cit. on p. 138).

[Kiy99] Kiyokawa, Kiyoshi; Takemura, Haruo and Yokoya, Naokazu: “A Collaboration Support Technique by Integrating a Shared Virtual Reality and a Shared Augmented Reality”. In: *IEEE SMC’99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 99CH37028).* Vol. 6. IEEE. 1999, pp. 48–53 (cit. on p. 28).

[Kni18] Knierim, Pascal; Schwind, Valentin; Feit, Anna Maria; Nieuwenhuizen, Florian and Henze, Niels: “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. Montreal QC, Canada: Association for Computing Machinery, 2018, pp. 1–9. doi: 10.1145/3173574.3173919. url: https://doi.org/10.1145/3173574.3173919 (cit. on pp. 17, 39, 40, 45, 46).

[Koe91] Koenderink, Jan J and Van Doorn, Andrea J: “Affine structure from motion”. In: *JOSA A* 8.2 (1991), pp. 377–385 (cit. on p. 86).

[Koh09] Kohli, Luv: “Exploiting Perceptual Illusions to Enhance Passive Haptics”. In: *IEEE VR Workshop on Perceptual Illusions in Virtual Environments.* Citeseer. 2009, pp. 22–24 (cit. on pp. 31, 115).

[Koh10] Kohli, L.: “Redirected touching: Warping space to remap passive haptics”. In: *2010 IEEE Symposium on 3D User Interfaces (3DUI).* Mar. 2010, pp. 129–130. doi: 10.1109/3DUI.2010.5444703 (cit. on pp. 31, 115).

[Koh12] Kohli, Luv; Whitton, Mary C and Brooks, Frederick P: “Redirected Touching: The Effect of Warping Space on Task Performance”. In: *2012 IEEE Symposium on 3D User Interfaces (3DUI).* IEEE. 2012, pp. 105–112 (cit. on p. 31).

[Koh13] Kohli, Luv; Whitton, Mary C and Brooks, Frederick P: “Redirected Touching: Training and Adaptation in Warped Virtual Spaces”. In: *2013 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE. 2013, pp. 79–86 (cit. on p. 31).

[Köl07] Költringer, Thomas; Isokoski, Poika and Grechenig, Thomas: “TwoStick: Writing with a game controller”. In: *Proceedings of Graphics interface 2007*. ACM. 2007, pp. 103–110 (cit. on pp. 13, 39, 40).

[Kol18] Kolkmeier, Jan; Harmsen, Emiel; Giesselink, Sander; Reidsma, Dennis; Theune, Mariët and Heylen, Dirk: “With a Little Help from a Holographic Friend: The OpenIMPRESS Mixed Reality Telepresence Toolkit for Remote Collaboration Systems”. In: *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*. ACM. 2018, p. 26 (cit. on pp. 3, 138).

[Kom17] Komiyama, Ryohei; Miyaki, Takashi and Rekimoto, Jun: “JackIn Space: Designing a Seamless Transition Between First and Third Person View for Effective Telepresence Collaborations”. In: *Proceedings of the 8th Augmented Human International Conference*. ACM. 2017, p. 14 (cit. on p. 28).

[Kon03] Konyo, M.; Akazawa, K.; Tadokoro, S. and Takamori, T.: “Tactile feel display for virtual active touch”. In: *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No.03CH37453)*. Vol. 4. 2003, 3744–3750 vol.3. doi: 10.1109/IROS.2003.1249737 (cit. on p. 16).

[Kos08] Koskinen, Emilia; Kaaresoja, Topi and Laitinen, Pauli: “Feel-Good Touch: Finding the Most Pleasant Tactile Feedback for a Mobile Touch Screen Button”. In: *Proceedings of the 10th International Conference on Multimodal Interfaces*. ICMI ’08. Chania, Crete, Greece: Association for Computing Machinery, 2008, pp. 297–304. doi: 10.1145/1452392.1452453 (cit. on p. 15).

[Kra06] Kranstedt, Alfred; Lücking, Andy; Pfeiffer, Thies; Rieser, Hannes and Staudacher, Marc: "Measuring and reconstructing pointing in visual contexts". In: *Proceedings of the brandial* (2006), pp. 82–89 (cit. on p. 27).

[Kro05] Krohn, A.; Beigl, M.; Hazas, M. and Gellersen, H.-W.: "Using Fine-Grained Infrared Positioning to Support the Surface-Based Activities of Mobile Users". In: *25th IEEE International Conference on Distributed Computing Systems Workshops.* 25th IEEE International Conference on Distributed Computing Systems Workshops. June 2005, pp. 463–468. doi: 10.1109/ICDCSW.2005.139 (cit. on p. 24).

[Kun20] Kunert, A.; Weissker, T.; Froehlich, B. and Kulik, A.: "Multi-Window 3D Interaction for Collaborative Virtual Reality". In: *IEEE Transactions on Visualization and Computer Graphics* 26.11 (2020), pp. 3271–3284. doi: 10.1109/TVCG.2019.2914677 (cit. on p. 30).

[Kur04] Kurata, Takeshi; Sakata, Nobuchika; Kourogi, Masakatsu; Kuzuoka, Hideaki and Billinghurst, Mark: "Remote Collaboration Using a Shoulder-Worn Active Camera/Laser". In: *Eighth International Symposium on Wearable Computers.* Vol. 1. IEEE. 2004, pp. 62–69 (cit. on p. 27).

[Kur14] Kurz, D.: "Thermal touch: Thermography-enabled everywhere touch interfaces for mobile augmented reality applications". In: *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR).* 2014, pp. 9–16. doi: 10.1109/ISMAR.2014.6948403 (cit. on p. 19).

[Kyu09] Kyung, K. -U. and Lee, J. -Y.: "Ubi-Pen: A Haptic Interface with Texture and Vibrotactile Display". In: *IEEE Computer Graphics and Applications* 29.1 (2009), pp. 56–64. doi: 10.1109/MCG.2009.17 (cit. on p. 16).

[Lac17] Lacoche, Jérémy; Pallamin, Nico; Boggini, Thomas and Royan, Jérôme: "Collaborators Awareness for User Cohabitation in Co-Located Collaborative Virtual Environments". In: *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*. VRST '17. Gothenburg, Sweden: Association for Computing Machinery, 2017. DOI: 10.1145/3139131.3139142. URL: https://doi.org/10.1145/3139131.3139142 (cit. on pp. 25, 26).

[Lan04] LANG, Stefan: "Skript Zur Vorlesung Computerintensive Verfahren in Der Statistik". In: (2004). URL: https://www.uibk.ac.at/statistics/personal/lang/publications/compstat_aktuell.pdf (visited on 02/28/2020) (cit. on p. 294).

[Lan18] LANGBEHN, Eike; HARTING, Eva and STEINICKE, Frank: "Shadow-avatars: A visualization method to avoid collisions of physically co-located users in room-scale VR". In: *IEEE Workshop on Everyday Virtual Reality*. 2018 (cit. on p. 26).

[Lap14] LAPUT, Gierad; XIAO, Robert; CHEN, Xiang 'Anthony'; HUDSON, Scott E. and HARRISON, Chris: "Skin Buttons: Cheap, Small, Low-Powered and Clickable Fixed-Icon Laser Projectors". In: *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*. UIST '14. Honolulu, Hawaii, USA: Association for Computing Machinery, 2014, pp. 389–394. DOI: 10.1145/2642918.2647356. URL: https://doi.org/10.1145/2642918.2647356 (cit. on p. 16).

[Lar03] LARIMER, Daniel and BOWMAN, Doug A: "VEWL: A Framework for Building a Windowing Interface in a Virtual Environment." In: *INTERACT*. 2003 (cit. on pp. 14, 16, 20).

[LaV11] LAVIOLA, Joseph J and KEEFE, Daniel F: "3D spatial interaction: applications for art, design, and science". In: *ACM Siggraph 2011 Courses*. ACM. 2011, p. 1 (cit. on p. 12).

[Le 15] LE CHÉNÉCHAL, Morgan; DUVAL, Thierry; GOURANTON, Valérie; ROYAN, Jérôme and ARNALDI, Bruno: "The Stretchable Arms for Collaborative Remote Guiding". In: *Proceedings of International Conference on Artificial Reality and Telexistence Eurographics Symposium on Virtual Environments*. Kyoto, Japan, 2015 (cit. on p. 27).

[Le 16a] Le Chénéchal, Morgan; Duval, Thierry; Gouranton, Valérie; Royan, Jérôme and Arnaldi, Bruno: “Vishnu: Virtual Immersive Support for HelpiNg Users - An Interaction Paradigm for Collaborative Remote Guiding in Mixed Reality”. In: *2016 IEEE Third VR International Workshop on Collaborative Virtual Environments (3DCVE).* IEEE. 2016, pp. 9–12 (cit. on pp. 27, 137).

[Le 16b] Le Chénéchal, Morgan; Lacoche, Jérémy; Royan, Jérôme; Duval, Thierry; Gouranton, Valérie and Arnaldi, Bruno: “When the Giant meets the Ant An Asymmetric Approach for Collaborative and Concurrent Object Manipulation in a Multi-Scale Environment”. In: *2016 IEEE Third VR International Workshop on Collaborative Virtual Environments (3DCVE).* IEEE. 2016, pp. 18–22 (cit. on p. 3).

[Lee11] Lee, Sangyoon and Hua, Hong: “Effects of viewing conditions and rotation methods in a collaborative tabletop AR environment”. In: *IEEE transactions on visualization and computer graphics* 17.9 (2011), pp. 1245–1258 (cit. on p. 32).

[Lee13] Lee, Jinha; Olwal, Alex; Ishii, Hiroshi and Boulanger, Cati: “SpaceTop: Integrating 2D and Spatial 3D Interactions in a See-through Desktop Environment”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* CHI ’13. Paris, France: Association for Computing Machinery, 2013, pp. 189–192. doi: 10.1145/2470654.2470680. url: https://doi.org/10.1145/2470654.2470680 (cit. on p. 17).

[Lee16a] Lee, Jihyun; Kim, Byungmoon; Suh, Bongwon and Koh, Eunyee: “Exploring the Front Touch Interface for Virtual Reality Headsets”. In: *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems.* ACM. 2016, pp. 2585–2591 (cit. on p. 13).

[Lee16b] Lee, Jihyun; Kim, Byungmoon; Suh, Bongwon and Koh, Eunyee: “Exploring the Front Touch Interface for Virtual Reality Headsets”. In: *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems.* CHI EA ’16.

San Jose, California, USA: Association for Computing Machinery, 2016, pp. 2585–2591. DOI: 10.1145/2851581.2892344. URL: https://doi.org/10.1145/2851581.2892344 (cit. on p. 15).

[Lee17] LEE, Gun A; TEO, Theophilus; KIM, Seungwon and BILLINGHURST, Mark: “Mixed Reality Collaboration Through Sharing a Live Panorama”. In: *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications*. ACM. 2017, p. 14 (cit. on p. 27).

[Lee18] LEE, Gun A; TEO, Theophilus; KIM, Seungwon and BILLINGHURST, Mark: “A User Study on MR Remote Collaboration Ising Live 360 Video”. In: *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE. 2018, pp. 153–164 (cit. on p. 27).

[Lee20] LEE, G.; KANG, H.; LEE, J. and HAN, J.: “A User Study on View-sharing Techniques for One-to-Many Mixed Reality Collaborations”. In: *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 2020, pp. 343–352. DOI: 10.1109/VR46266.2020.00054 (cit. on p. 28).

[Lei15] LEIGH, Sang-won; SCHOESSLER, Philipp; HEIBECK, Felix; MAES, Pattie and ISHII, Hiroshi: “THAW: Tangible Interaction with See-Through Augmentation for Smartphones on Computer Screens”. In: *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction*. TEI ’15. Stanford, California, USA: Association for Computing Machinery, 2015, pp. 89–96. DOI: 10.1145/2677199.2680584. URL: https://doi.org/10.1145/2677199.2680584 (cit. on pp. 16–18).

[Lep17] LEPOURAS, Georgios: “Comparing methods for numerical input in immersive virtual environments”. In: *Virtual Reality* (2017), pp. 1–15 (cit. on p. 13).

[Lia13] LIANG, Hai-Ning; WILLIAMS, Cary; SEMEGEN, Myron; STUERZLINGER, Wolfgang and IRANI, Pourang: “An Investigation of Suitable Interactions for 3D Manipulation of Distant Objects

Through a Mobile Device”. In: *International Journal of Innovative Computing, Information and Control* 9.12 (2013), pp. 4737–4752 (cit. on p. 15).

[Lin13] Lin, Qiufeng; Rieser, John J. and Bodenheimer, Bobby: “Stepping off a Ledge in an HMD-Based Immersive Virtual Environment”. In: *Proceedings of the ACM Symposium on Applied Perception*. SAP ’13. Dublin, Ireland: Association for Computing Machinery, 2013, pp. 107–110. doi: 10.1145/2492494.2492511. url: https://doi.org/10.1145/2492494.2492511 (cit. on p. 13).

[Lin17] Lin, Jia-Wei; Han, Ping-Hsuan; Lee, Jiun-Yu; Chen, Yang-Sheng; Chang, Ting-Wei; Chen, Kuan-Wen and Hung, Yi-Ping: “Visualizing the keyboard in virtual reality for enhancing immersive experience”. In: *ACM SIGGRAPH 2017 Posters*. ACM. 2017, p. 35 (cit. on pp. 13, 39, 40).

[Lin99] Lindeman, R. W.; Sibert, J. L. and Hahn, J. K.: “Hand-held windows: towards effective 2D interaction in immersive virtual environments”. In: *Proceedings IEEE Virtual Reality (Cat. No. 99CB36316)*. Mar. 1999, pp. 205–212. doi: 10.1109/VR.1999.756952 (cit. on pp. 12, 14, 18, 20).

[Lis14] Lissermann, Roman; Huber, Jochen; Schmitz, Martin; Steimle, Jürgen and Mühlhäuser, Max: “Permulin: mixed-focus collaboration on multi-view tabletops”. In: *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. ACM. 2014, pp. 3191–3200 (cit. on p. 29).

[Lit65] Little, Kenneth B: “Personal Space”. In: *Journal of Experimental Social Psychology* (1965) (cit. on p. 30).

[Liu18] Liu, Y.; Dong, H.; Zhang, L. and Saddik, A. E.: “Technical Evaluation of HoloLens for Multimedia: A First Look”. In: *IEEE MultiMedia* 25.4 (Oct. 2018), pp. 8–18. doi: 10.1109/MMUL.2018.2873473 (cit. on p. 24).

[Llo07] Lloyd, Donna M.: "Spatial limits on referred touch to an alien limb may reflect boundaries of visuo-tactile peripersonal space surrounding the hand". In: *Brain and Cognition* 64.1 (2007), pp. 104–109. doi: https://doi.org/10.1016/j.bandc.2006.09.013. url: http://www.sciencedirect.com/science/article/pii/S0278262606002119 (cit. on p. 31).

[Loo07] Looser, J.; Grasset, R. and Billinghurst, M.: "A 3D Flexible and Tangible Magic Lens in Augmented Reality". In: *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*. 2007, pp. 51–54. doi: 10.1109/ISMAR.2007.4538825 (cit. on p. 16).

[Lóp16] López, D.; Oehlberg, L.; Doger, C. and Isenberg, T.: "Towards An Understanding of Mobile Touch Navigation in a Stereoscopic Viewing Environment for 3D Data Exploration". In: *IEEE Transactions on Visualization and Computer Graphics* 22.5 (2016), pp. 1616–1629. doi: 10.1109/TVCG.2015.2440233 (cit. on pp. 15, 17, 18).

[Low04] Lowe, David G.: "Distinctive Image Features from Scale-Invariant Keypoints". In: *International Journal of Computer Vision* 60 (2004), pp. 91–110. url: https://doi.org/10.1023/B:VISI.0000029664.99615.94 (cit. on p. 86).

[Lux16] Luxenburger, Andreas; Prange, Alexander; Moniri, Mohammad Mehdi and Sonntag, Daniel: "MedicaLVR: towards medical remote collaboration using virtual reality". In: *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*. ACM. 2016, pp. 321–324 (cit. on p. 2).

[Mac02] MacKenzie, I. Scott: "KSPC (Keystrokes per Character) as a Characteristic of Text Entry Techniques". In: *Human Computer Interaction with Mobile Devices*. Ed. by Paternò, Fabio. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002, pp. 195–210 (cit. on p. 39).

[Man18] Mandalika, Veera Bhadra Harish; Chernoglazov, Alexander I; Billinghurst, Mark; Bartneck, Christoph; Hurrell, Michael A; De Ruiter, Niels; Butler, Anthony PH and Butler, Philip H: "A hybrid 2D/3D user interface for radiological diagnosis". In: *Journal of digital imaging* 31.1 (Feb. 2018), pp. 56–73. url: https://doi.org/10.1007/s10278-017-0002-6 (cit. on p. 19).

[Man47] Mann, H. B. and Whitney, D. R.: "On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other". In: *The Annals of Mathematical Statistics* 18.1 (1947), pp. 50–60. doi: 10.1214/aoms/1177730491 (cit. on p. 296).

[Man98] Mankoff, Jennifer and Abowd, Gregory D: "Cirrin: a word-level unistroke keyboard for pen input". In: *Proceedings of the 11th annual ACM symposium on User interface software and technology.* ACM. 1998, pp. 213–214 (cit. on p. 13).

[Mar10] Martinez, Diego; Molina, Jose P; Garcia, Arturo S; Martinez, Jonatan and Gonzalez, Pascual: "AFreeCA: Extending the spatial model of interaction". In: *Cyberworlds (CW), 2010 International Conference on.* IEEE. 2010, pp. 17–24 (cit. on p. 29).

[Mar12] Markov-Vetter, Daniela; Moll, Eckard and Staadt, Oliver: "Evaluation of 3D Selection Tasks in Parabolic Flight Conditions: Pointing Task in Augmented Reality User Interfaces". In: *Proceedings of the 11th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry.* VRCAI '12. Singapore, Singapore: Association for Computing Machinery, 2012, pp. 287–294. doi: 10.1145/2407516.2407583 (cit. on p. 15).

[Mar14a] Marks, Stefan; Estevez, Javier E and Connor, Andy M: "Towards the Holodeck: fully immersive virtual reality visualisation of scientific and engineering data". In: *Proceedings of the 29th International Conference on Image and Vision Computing New Zealand.* ACM. 2014, pp. 42–47 (cit. on p. 2).

[Mar14b] Marzo, Asier; Bossavit, Benoit and Hachet, Martin: “Combining Multi-Touch Input and Device Movement for 3D Manipulations in Mobile Augmented Reality Environments”. In: *Proceedings of the 2nd ACM Symposium on Spatial User Interaction.* SUI ’14. Honolulu, Hawaii, USA: Association for Computing Machinery, 2014, pp. 13–16. doi: 10.1145/2659766.2659775. url: https://doi.org/10.1145/2659766.2659775 (cit. on pp. 15, 17, 18).

[Mar18] Marwecki, Sebastian; Brehm, Maximilian; Wagner, Lukas; Cheng, Lung-Pan; Mueller, Florian ’Floyd’ and Baudisch, Patrick: “VirtualSpace - Overloading Physical Space with Multiple Virtual Reality Users”. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.* CHI ’18. Montreal QC, Canada: Association for Computing Machinery, 2018, pp. 1–10. doi: 10.1145/3173574.3173815. url: https://doi.org/10.1145/3173574.3173815 (cit. on p. 25).

[Mar19] Marsh, Ronald: “Survey on depth perception in head mounted displays: distance estimation in virtual reality, augmented reality, and mixed reality”. English. In: *IET Image Processing* 13 (5 Apr. 2019), 707–712(5). url: https://digital-library.theiet.org/content/journals/10.1049/iet-ipr.2018.5920 (cit. on p. 10).

[Mas94] Massie, Thomas H and Salisbury, J Kenneth: “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.* Vol. 55. 1. Chicago, IL. 1994, pp. 295–300 (cit. on p. 16).

[Mat04] Matsushita, Mitsunori; Iida, Makoto; Ohguro, Takeshi; Shirai, Yoshinari; Kakehi, Yasuaki and Naemura, Takeshi: “Lumisight table: a face-to-face collaboration support system that optimizes direction of projected information to each stakeholder”. In: *Proceedings of the 2004 ACM conference on Computer supported cooperative work.* ACM. 2004, pp. 274–283 (cit. on p. 29).

[Mat15] Matsukage, Ryo; Iwai, Daisuke and Sato, Kosuke: "Consistent desktop sharing based on document coordinate system for face-to-face online meeting". In: *SIGGRAPH Asia 2015 Emerging Technologies.* ACM. 2015, p. 6 (cit. on p. 26).

[Mat17] Matsubara, T.; Mori, N.; Niikura, T. and Tano, S.: "Touch detection method for non-display surface using multiple shadows of finger". In: *2017 IEEE 6th Global Conference on Consumer Electronics (GCCE).* 2017, pp. 1–5. doi: 10.1109/GCCE.2017.8229364 (cit. on p. 18).

[Maz96] Mazuryk, Tomasz and Gervautz, Michael: "Virtual reality-history, applications, technology and future". In: (1996) (cit. on p. 10).

[McG15] McGill, Mark; Boland, Daniel; Murray-Smith, Roderick and Brewster, Stephen: "A dose of reality: Overcoming usability challenges in vr head-mounted displays". In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems.* ACM. 2015, pp. 2143–2152 (cit. on pp. 39, 40, 45).

[McM11] McManus, Erin A.; Bodenheimer, Bobby; Streuber, Stephan; Rosa, Stephan de la; Bülthoff, Heinrich H. and Mohler, Betty J.: "The Influence of Avatar (Self and Character) Animations on Distance Estimation, Object Interaction and Locomotion in Immersive Virtual Environments". In: *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization.* APGV '11. Toulouse, France: Association for Computing Machinery, 2011, pp. 37–44. doi: 10.1145/2077451.2077458. url: https://doi.org/10.1145/2077451.2077458 (cit. on p. 13).

[Med13] Medeiros, Daniel; Teixeira, Lucas; Carvalho, Felipe; Santos, Ismael and Raposo, Alberto: "A Tablet-Based 3D Interaction Tool for Virtual Engineering Environments". In: *Proceedings of the 12th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry.* VRCAI '13. Hong Kong, Hong Kong: Association for Computing Machinery, 2013, pp. 211–218. DOI: 10.1145/2534329.2534349 (cit. on pp. 17, 18).

[Men20] MENZNER, T.; GESSLEIN, T.; OTTE, A. and GRUBERT, J.: “Above Surface Interaction for Multiscale Navigation in Mobile Virtual Reality”. In: *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR).* 2020, pp. 372–381. DOI: 10.1109/VR46266.2020.00057 (cit. on p. 67).

[Mil91] MILLS, David L: “Internet time synchronization: the network time protocol”. In: *IEEE Transactions on communications* 39.10 (1991), pp. 1482–1493 (cit. on p. 82).

[Mil95] MILGRAM, Paul; TAKEMURA, Haruo; UTSUMI, Akira and KISHINO, Fumio: “Augmented reality: a class of displays on the reality-virtuality continuum”. In: *Telemanipulator and Telepresence Technologies.* Ed. by DAS, Hari. Vol. 2351. International Society for Optics and Photonics. SPIE, 1995, pp. 282–292. DOI: 10.1117/12.197321. URL: https://doi.org/10.1117/12.197321 (cit. on pp. 67, 179).

[Min95] MINE, Mark: “ISAAC: A virtual environment tool for the interactive construction of virtual worlds”. In: *UNC Chapel Hill Computer Science Technical Report TR95-020* (1995) (cit. on p. 15).

[Min96] MINE, Mark: “Working in a virtual world: Interaction techniques used in the chapel hill immersive modeling program”. In: (1996) (cit. on pp. 14, 15).

[Min97a] MINE, Mark R; BROOKS JR, Frederick P and SEQUIN, Carlo H: “Moving objects in space: exploiting proprioception in virtual-environment interaction”. In: *Proceedings of the 24th annual conference on Computer graphics and interactive techniques.* ACM Press/Addison-Wesley Publishing Co. 1997, pp. 19–26 (cit. on pp. 11, 14).

[Min97b] MINE, Mark R.; BROOKS, Frederick P. and SEQUIN, Carlo H.: “Moving Objects in Space: Exploiting Proprioception in Virtual-Environment Interaction”. In: *Proceedings of the 24th Annual*

Conference on Computer Graphics and Interactive Techniques. SIGGRAPH '97. USA: ACM Press/Addison-Wesley Publishing Co., 1997, pp. 19–26. DOI: 10.1145/258734.258747 (cit. on p. 20).

[Miy18] Miyazaki, M. and Komuro, T.: "Extended Workspace Using a Smartphone with a Depth Camera". In: *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct).* 2018, pp. 115–116. DOI: 10.1109/ISMAR-Adjunct.2018.00046 (cit. on p. 52).

[Moh08] Mohler, Betty J.; Bülthoff, Heinrich H.; Thompson, William B. and Creem-Regehr, Sarah H.: "A Full-Body Avatar Improves Egocentric Distance Judgments in an Immersive Virtual Environment". In: *Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization.* APGV '08. Los Angeles, California: Association for Computing Machinery, 2008, p. 194. DOI: 10.1145/1394281.1394323. URL: https://doi.org/10.1145/1394281.1394323 (cit. on p. 13).

[Moh10] Mohler, Betty J; Creem-Regehr, Sarah H; Thompson, William B and Bülthoff, Heinrich H: "The Effect of Viewing a Self-Avatar on Distance Judgments in an HMD-Based Virtual Environment". In: *Presence: Teleoperators and Virtual Environments* 19.3 (2010), pp. 230–242. DOI: 10.1162/pres.19.3.230. URL: https://doi.org/10.1162/pres.19.3.230 (cit. on pp. 11, 13).

[Möl11] Möllers, Max and Borchers, Jan: "TaPS widgets: interacting with tangible private spaces". In: *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces.* ACM. 2011, pp. 75–78 (cit. on p. 29).

[Mon11] Montoya, Mitzi M.; Massey, Anne P. and Lockwood, Nicholas S.: "3D Collaborative Virtual Environments: Exploring the Link between Collaborative Behaviors and Team Performance". In: *Decision Sciences* 42.2 (2011), pp. 451–476. DOI: 10.1111/j.1540-5915.2011.00318.x (cit. on pp. 3, 148).

[Mon17] Montano Murillo, Roberto A; Subramanian, Sriram and Martinez Plasencia, Diego: “Erg-O: Ergonomic Optimization of Immersive Virtual Environments”. In: *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology.* ACM. 2017, pp. 759–771 (cit. on p. 31).

[Moo07] Moore, Robert J; Ducheneaut, Nicolas and Nickell, Eric: “Doing virtually nothing: Awareness and accountability in massively multiplayer online worlds”. In: *Computer Supported Cooperative Work (CSCW)* 16.3 (2007), pp. 265–305 (cit. on p. 32).

[Mor12] Mori, Masahiro; MacDorman, Karl F and Kageki, Norri: “The uncanny valley [from the field]”. In: *IEEE Robotics & Automation Magazine* 19.2 (2012), pp. 98–100 (cit. on p. 104).

[Mos13] Mossel, Annette; Venditti, Benjamin and Kaufmann, Hannes: “3DTouch and HOMER-S: Intuitive Manipulation Techniques for One-Handed Handheld Augmented Reality”. In: *Proceedings of the Virtual Reality International Conference: Laval Virtual.* VRIC ’13. Laval, France: Association for Computing Machinery, 2013. doi: 10.1145/2466816.2466829. url: https://doi.org/10.1145/2466816.2466829 (cit. on pp. 15, 17, 18).

[Mou17] Mousas, Christos: “Full-body locomotion reconstruction of virtual characters using a single inertial measurement unit”. In: *Sensors* 17.11 (2017), p. 2589. doi: 10.3390/s17112589. url: https://doi.org/10.3390/s17112589 (cit. on p. 103).

[Muj04] Mujber, T.S.; Szecsi, T. and Hashmi, M.S.J.: “Virtual Reality Applications in Manufacturing Process Simulation”. In: *Journal of Materials Processing Technology* 155-156 (Nov. 2004), pp. 1834–1838. doi: 10.1016/j.jmatprotec.2004.04.401 (cit. on p. 9).

[Mul98] Muller, J; Krapichler, Christian; Nguyen, Lam Son; Englmeier, K Hans and Lang, Manfred: “Speech interaction in virtual reality”. In: *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on.* Vol. 6. IEEE. 1998, pp. 3757–3760 (cit. on p. 14).

[Mur12] Murugappan, Sundar; Vinayak; Elmqvist, Niklas and Ramani, Karthik: “Extended Multitouch: Recovering Touch Posture and Differentiating Users Using a Depth Camera”. In: *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*. UIST ’12. Cambridge, Massachusetts, USA: Association for Computing Machinery, 2012, pp. 487–496. doi: 10.1145/2380116.2380177. url: https://doi.org/10.1145/2380116.2380177 (cit. on p. 18).

[Nag15] Nagai, Shohei; Kasahara, Shunichi and Rekimoto, Jun: “Livesphere: Sharing the Surrounding Visual Environment for Immersive Experience in Remote Collaboration”. In: *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction*. ACM. 2015, pp. 113–116 (cit. on pp. 29, 30).

[Nal06] Nalbant, G and Bostan, B: “Interaction in virtual reality”. In: *4th international symposium of interactive medical design (ISIMD)*. 2006 (cit. on p. 9).

[Ngu16] Nguyen, Minh-Tu; Nguyen, Hai-Khanh; Vo-Lam, Khanh-Duy; Nguyen, Xuan-Gieng and Tran, Minh-Triet: “Applying virtual reality in city planning”. In: *International Conference on Virtual, Augmented and Mixed Reality*. Springer. 2016, pp. 724–735 (cit. on p. 2).

[Ngu17] Nguyen, Cuong; DiVerdi, Stephen; Hertzmann, Aaron and Liu, Feng: “Vremiere: In-Headset Virtual Reality Video Editing”. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. CHI ’17. Denver, Colorado, USA: Association for Computing Machinery, 2017, pp. 5428–5438. doi: 10.1145/3025453.3025675. url: https://doi.org/10.1145/3025453.3025675 (cit. on p. 19).

[Ni11] Ni, Tao; Bowman, Doug and North, Chris: “AirStroke: bringing unistroke text entry to freehand gesture interfaces”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM. 2011, pp. 2473–2476 (cit. on pp. 14, 39, 40).

[Nie17a] Niehorster, Diederick C.; Li, Li and Lappe, Markus: “The Accuracy and Precision of Position and Orientation Tracking in the HTC Vive Virtual Reality System for Scientific Research”. In: *i-Perception* 8.3 (2017). doi: 10.1177/2041669517708205 (cit. on p. 54).

[Nie17b] Niehorster, Diederick C.; Li, Li and Lappe, Markus: “The Accuracy and Precision of Position and Orientation Tracking in the HTC Vive Virtual Reality System for Scientific Research”. In: *i-Perception* 8.3 (June 2017), p. 204166951770820. doi: 10.1177/2041669517708205 (cit. on p. 87).

[Nie17c] Nierula, Birgit; Martini, Matteo; Matamala-Gomez, Marta; Slater, Mel and Sanchez-Vives, Maria V.: “Seeing an Embodied Virtual Hand is Analgesic Contingent on Colocation”. In: *The Journal of Pain* 18.6 (2017), pp. 645–655. doi: https://doi.org/10.1016/j.jpain.2017.01.003. url: http://www.sciencedirect.com/science/article/pii/S1526590017300172 (cit. on p. 31).

[Nii16] Niikura, Takehiro; Matsubara, Takashi and Mori, Naoki: “Touch Detection System for Various Surfaces Using Shadow of Finger”. In: *Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces*. ISS ’16. Niagara Falls, Ontario, Canada: Association for Computing Machinery, 2016, pp. 337–342. doi: 10.1145/2992154.2996777. url: https://doi.org/10.1145/2992154.2996777 (cit. on p. 18).

[Nit18] Nittala, Aditya Shekhar; Withana, Anusha; Pourjafarian, Narjes and Steimle, Jürgen: “Multi-Touch Skin: A Thin and Flexible Multi-Touch Sensor for On-Skin Input”. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI ’18. Montreal QC, Canada: Association for Computing Machinery, 2018, pp. 1–12. doi: 10.1145/3173574.3173607. url: https://doi.org/10.1145/3173574.3173607 (cit. on p. 16).

[Nte17] Ntelidakis, Antonios; Zabulis, Xenophon; Grammenos, Dimitris and Koutlemanis, Panagiotis: “Touch Detection for Planar

Interactive Displays Based on Lateral Depth Views". In: *Multimedia Tools Appl.* 76.10 (May 2017), pp. 12683–12707. DOI: 10.1007/s11042-016-3695-5. URL: https://doi.org/10.1007/s11042-016-3695-5 (cit. on p. 19).

[Oda15] ODA, Ohan; ELVEZIO, Carmine; SUKAN, Mengu; FEINER, Steven and TVERSKY, Barbara: "Virtual Replicas for Remote Assistance in Virtual and Augmented Reality". In: *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology.* ACM. 2015, pp. 405–415 (cit. on p. 3).

[Oh06] OH, J. and HUA, H.: "User evaluations on form factors of tangible magic lenses". In: *2006 IEEE/ACM International Symposium on Mixed and Augmented Reality.* 2006, pp. 23–32. DOI: 10.1109/ISMAR.2006.297790 (cit. on p. 16).

[Oh18] OH, Catherine S.; BAILENSON, Jeremy N. and WELCH, Gregory F.: "A Systematic Review of Social Presence: Definition, Antecedents, and Implications". In: *Frontiers in Robotics and AI* 5 (2018), p. 114. DOI: 10.3389/frobt.2018.00114. URL: https://www.frontiersin.org/article/10.3389/frobt.2018.00114 (cit. on p. 9).

[Oh20a] OH, J. Y.; PARK, J. and PARK, J.: "FingerTouch: Touch Interaction Using a Fingernail-Mounted Sensor on a Head-Mounted Display for Augmented Reality". In: *IEEE Access* 8 (2020), pp. 101192–101208. DOI: 10.1109/ACCESS.2020.2997972 (cit. on p. 19).

[Oh20b] OH, Seo Young; YOON, Boram; KIM, Hyung-il and WOO, Woontack: "Finger Contact in Gesture Interaction Improves Time-Domain Input Accuracy in HMD-Based Augmented Reality". In: *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems.* CHI EA '20. Honolulu, HI, USA: Association for Computing Machinery, 2020, pp. 1–8. DOI: 10.1145/3334480.3383098. URL: https://doi.org/10.1145/3334480.3383098 (cit. on p. 15).

[Oli18] Olivier, A.; Bruneau, J.; Kulpa, R. and Pettré, J.: "Walking with Virtual People: Evaluation of Locomotion Interfaces in Dynamic Environments". In: *IEEE Transactions on Visualization and Computer Graphics* 24.7 (2018), pp. 2251–2263. doi: 10.1109/TVCG.2017.2714665 (cit. on pp. 23, 25).

[Ols11] Olson, Edwin: "AprilTag: A robust and flexible visual fiducial system". In: *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA).* IEEE, May 2011, pp. 3400–3407 (cit. on p. 86).

[Osh10] Oshima, Ryo and Kakehi, Yasuaki: "Bounsight table: a view-dependent display with a single front projection". In: *ACM International Conference on Interactive Tabletops and Surfaces.* ACM. 2010, pp. 281–282 (cit. on p. 29).

[Otm11] Otmane, Samir; Domingues, Christophe; Davesne, Frédéric and Mallem, Malik: "Collaborative 3D Interaction in Virtual Environments: A Workflow-Based Approach". In: *Virtual Reality Intech International* (2011), p. 49 (cit. on p. 3).

[Ott02] Ott, David and Dillenbourg, Pierre: "Proximity and View Awareness to Reduce Referential Ambiguity in a Shared 3D Virtual Environment". In: *Proceedings of the Conference on Computer Support for Collaborative Learning: Foundations for a CSCL Community.* CSCL '02. Boulder, Colorado: International Society of the Learning Sciences, 2002, pp. 603–604 (cit. on p. 29).

[Ott06] Otto, Oliver; Roberts, Dave and Wolff, Robin: "A Review on Effective Closely-Coupled Collaboration Using Immersive CVE's". In: *Proceedings of the 2006 ACM International Conference on Virtual Reality Continuum and Its Applications.* VRCIA '06. Hong Kong, China: Association for Computing Machinery, 2006, pp. 145–154. doi: 10.1145/1128923.1128947 (cit. on p. 3).

[Ott19] Otte, A.; Schneider, D.; Menzner, T.; Gesslein, T.; Gagel, P. and Grubert, J.: "Evaluating Text Entry in Virtual Reality using a Touch-sensitive Physical Keyboard". In: *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct*

(ISMAR-Adjunct). 2019, pp. 387–392. DOI: 10.1109/ISMAR-Adjunct.2019.000-4 (cit. on p. 17).

[Pad16] PADRAO, Gonçalo; GONZALEZ-FRANCO, Mar; SANCHEZ-VIVES, Maria V.; SLATER, Mel and RODRIGUEZ-FORNELLS, Antoni: "Violating body movement semantics: Neural signatures of self-generated and external-generated errors". In: *NeuroImage* 124 (2016), pp. 147–156. DOI: https://doi.org/10.1016/j.neuroimage.2015.08.022. URL: http://www.sciencedirect.com/science/article/pii/S1053811915007314 (cit. on p. 13).

[Par00] PARK, Kyoung S; KAPOOR, Abhinav and LEIGH, Jason: "Lessons Learned from Employing Multiple Perspectives in a Collaborative Virtual Environment for Visualizing Scientific Data". In: *Proceedings of the third International Conference on Collaborative Virtual Environments.* ACM. 2000, pp. 73–82 (cit. on p. 3).

[Par06] PARHI, Pekka; KARLSON, Amy K. and BEDERSON, Benjamin B.: "Target Size Study for One-Handed Thumb Use on Small Touchscreen Devices". In: *Proceedings of the 8th Conference on Human-Computer Interaction with Mobile Devices and Services.* MobileHCI '06. Helsinki, Finland: Association for Computing Machinery, 2006, pp. 203–210. DOI: 10.1145/1152215.1152260 (cit. on p. 65).

[Par16] PARK, Ji Yeol; PARK, Jinwon; KIM, Kyumok; LEE, Jon-Ha and JUNG, Seung-Won: "A Development of Touch Sensing Using a Depth Camera–Projector System". In: *Advanced Multimedia and Ubiquitous Engineering.* Ed. by PARK, James J. (Jong Hyuk); JIN, Hai; JEONG, Young-Sik and KHAN, Muhammad Khurram. Singapore: Springer Singapore, 2016, pp. 175–180 (cit. on p. 18).

[Par17] PARZER, Patrick; SHARMA, Adwait; VOGL, Anita; STEIMLE, Jürgen; OLWAL, Alex and HALLER, Michael: "SmartSleeve: Real-Time Sensing of Surface and Deformation Gestures on Flexible, Interactive Textiles, Using a Hybrid Gesture Detection Pipeline". In: *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology.* UIST '17. Québec City, QC, Canada: Association for Computing Machinery, 2017, pp. 565–577. doi: 10.1145/3126594.3126652. url: https://doi.org/10.1145/3126594.3126652 (cit. on p. 16).

[Par19] Parizi, Farshid Salemi; Whitmire, Eric and Patel, Shwetak: "AuraRing: Precise Electromagnetic Finger Tracking". In: *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3.4 (Dec. 2019). doi: 10.1145/3369831. url: https://doi.org/10.1145/3369831 (cit. on p. 11).

[Pau10] Paul, Sharoda A and Reddy, Madhu C: "Understanding together: sensemaking in collaborative information seeking". In: *Proceedings of the 2010 ACM conference on Computer supported cooperative work*. ACM. 2010, pp. 321–330 (cit. on p. 22).

[Pea31] Pearson, Egon S.: "The Test of Significance for the Correlation Coefficient". In: *Journal of the American Statistical Association* 26.174 (1931), pp. 128–134. doi: 10.1080/01621459.1931.10503208. url: https://www.tandfonline.com/doi/abs/10.1080/01621459.1931.10503208 (cit. on p. 298).

[Peñ09] Peña, Jorge; Hancock, Jeffrey T. and Merola, Nicholas A.: "The Priming Effects of Avatars in Virtual Settings". In: *Communication Research* 36.6 (2009), pp. 838–856. doi: 10.1177/0093650209346802. url: https://doi.org/10.1177/0093650209346802 (cit. on p. 13).

[Per98] Perlin, Ken: "Quikwriting: continuous stylus-based text entry". In: *Proceedings of the 11th annual ACM symposium on User interface software and technology*. ACM. 1998, pp. 215–216 (cit. on p. 13).

[Pes12] Peschke, Joshua; Göbel, Fabian; Gründer, Thomas; Keck, Mandy; Kammer, Dietrich and Groh, Rainer: "DepthTouch: An Elastic Surface for Tangible Computing". In: *Proceedings of the International Working Conference on Advanced Visual Interfaces*. AVI '12. Capri Island, Italy: Association for Computing Machinery, 2012, pp. 770–771. doi: 10.1145/2254556.2254706. url: https://doi.org/10.1145/2254556.2254706 (cit. on p. 16).

[Pfe15] Pfeuffer, Ken; Alexander, Jason; Chong, Ming Ki; Zhang, Yanxia and Gellersen, Hans: “Gaze-Shifting: Direct-Indirect Input with Pen and Touch Modulated by Gaze”. In: *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. UIST ’15. Charlotte, NC, USA: Association for Computing Machinery, 2015, pp. 373–383. doi: 10.1145/2807442.2807460. url: https://doi.org/10.1145/2807442.2807460 (cit. on p. 14).

[Pfe16] Pfeuffer, Ken and Gellersen, Hans: “Gaze and Touch Interaction on Tablets”. In: *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. UIST ’16. Tokyo, Japan: Association for Computing Machinery, 2016, pp. 301–311. doi: 10.1145/2984511.2984514. url: https://doi.org/10.1145/2984511.2984514 (cit. on p. 14).

[Pfe17] Pfeuffer, Ken; Mayer, Benedikt; Mardanbegi, Diako and Gellersen, Hans: “Gaze + Pinch Interaction in Virtual Reality”. In: *Proceedings of the 5th Symposium on Spatial User Interaction*. SUI ’17. Brighton, United Kingdom: Association for Computing Machinery, 2017, pp. 99–108. doi: 10.1145/3131277.3132180. url: https://doi.org/10.1145/3131277.3132180 (cit. on p. 14).

[Piu17a] Piumsomboon, Thammathip; Day, Arindam; Ens, Barrett; Lee, Youngho; Lee, Gun and Billinghurst, Mark: “Exploring Enhancements for Remote Mixed Reality Collaboration”. In: *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications*. ACM. 2017, p. 16 (cit. on p. 28).

[Piu17b] Piumsomboon, Thammathip; Lee, Youngho; Lee, Gun and Billinghurst, Mark: “CoVAR: A Collaborative Virtual and Augmented Reality System for Remote Collaboration”. In: *SIGGRAPH Asia 2017 Emerging Technologies*. ACM. 2017, p. 3 (cit. on p. 28).

[Piu18] Piumsomboon, Thammathip; Lee, Gun A; Hart, Jonathon D; Ens, Barrett; Lindeman, Robert W; Thomas, Bruce H and Billinghurst, Mark: “Mini-Me: An Adaptive Avatar for Mixed

Reality Remote Collaboration”. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.* ACM. 2018, p. 46 (cit. on pp. 32, 115, 140, 175, 294).

[Piu19] Piumsomboon, Thammathip; Lee, Gun A; Irlitti, Andrew; Ens, Barrett; Thomas, Bruce H and Billinghurst, Mark: “On the Shoulder of the Giant: A Multi-Scale Mixed Reality Collaboration with 360 Video Sharing and Tangible Interaction”. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.* ACM. 2019, p. 228 (cit. on pp. 30, 32, 294).

[Pla20] Plasson, Carole; Cunin, Dominique; Laurillau, Yann and Nigay, Laurence: “3D Tabletop AR: A Comparison of Mid-Air, Touch and Touch+Mid-Air Interaction”. In: *Proceedings of the International Conference on Advanced Visual Interfaces.* AVI ’20. Salerno, Italy: Association for Computing Machinery, 2020. doi: 10.1145/3399715.3399836. url: https://doi.org/10.1145/3399715.3399836 (cit. on p. 15).

[Pod17] Podkosova, Iana and Kaufmann, Hannes: “Preventing imminent collisions between co-located users in HMD-based VR in non-shared scenarios”. In: *Proceedings of the 30th International Conference on Computer Animation and Social Agents, CASA.* 2017 (cit. on pp. 23, 25, 103).

[Pod18a] Podkosova, Iana and Kaufmann, Hannes: “Co-Presence and Proxemics in Shared Walkable Virtual Environments with Mixed Colocation”. In: *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology.* VRST ’18. Tokyo, Japan: Association for Computing Machinery, 2018. doi: 10 . 1145 / 3281505 . 3281523. url: https://doi.org/10.1145/3281505.3281523 (cit. on pp. 23, 25, 112).

[Pod18b] Podkosova, Iana and Kaufmann, Hannes: “Mutual Collision Avoidance during Walking in Real and Collaborative Virtual Environments”. In: *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games.* I3D ’18. Montreal, Quebec, Canada: Association for Computing Machinery, 2018.

DOI: 10.1145/3190834.3190845. URL: https://doi.org/10.1145/3190834.3190845 (cit. on p. 25).

[Pou16] Pouliquen-Lardy, Lauriane; Milleville-Pennel, Isabelle; Guillaume, François and Mars, Franck: “Remote collaboration in virtual reality: asymmetrical effects of task distribution on spatial processing and mental workload”. In: *Virtual Reality* 20.4 (2016), pp. 213–220 (cit. on pp. 29, 30).

[Pou19] Pourjafarian, Narjes; Withana, Anusha; Paradiso, Joseph A. and Steimle, Jürgen: “Multi-Touch Kit: A Do-It-Yourself Technique for Capacitive Multi-Touch Sensing Using a Commodity Microcontroller”. In: *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology.* UIST ’19. New Orleans, LA, USA: Association for Computing Machinery, 2019, pp. 1071–1083. DOI: 10.1145/3332165.3347895. URL: https://doi.org/10.1145/3332165.3347895 (cit. on p. 16).

[Pou96] Poupyrev, Ivan; Billinghurst, Mark; Weghorst, Suzanne and Ichikawa, Tadao: “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.* ACM. 1996, pp. 79–80 (cit. on pp. 31, 131).

[Pou98a] Poupyrev, I.; Ichikawa, T.; Weghorst, S. and Billinghurst, M.: “Egocentric Object Manipulation in Virtual Environments: Empirical Evaluation of Interaction Techniques”. In: *Computer Graphics Forum* 17.3 (1998), pp. 41–52. DOI: 10.1111/1467-8659.00252 (cit. on pp. 54, 105).

[Pou98b] Poupyrev, Ivan; Tomokazu, Numada and Weghorst, Suzanne: “Virtual Notepad: handwriting in immersive VR”. In: *Virtual Reality Annual International Symposium, 1998. Proceedings., IEEE 1998.* IEEE. 1998, pp. 126–132 (cit. on p. 13).

[Pow98] Powell, M. J. D.: “Direct Search Algorithms for Optimization Calculations”. In: *Acta Numerica* 7 (Jan. 1998), pp. 287–336. DOI: 10.1017/S0962492900002841 (cit. on p. 84).

[Prä14] Prätorius, Manuel; Valkov, Dimitar; Burgbacher, Ulrich and Hinrichs, Klaus: “DigiTap: An Eyes-Free VR/AR Symbolic Input Device”. In: *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology.* VRST ’14. Edinburgh, Scotland: Association for Computing Machinery, 2014, pp. 9–18. doi: 10.1145/2671015.2671029. url: https://doi.org/10.1145/2671015.2671029 (cit. on p. 19).

[Qin16] Qin, Zong; Chang, Yu-Cheng; Su, Yu-Jie; Huang, Yi-Pai and Shieh, Han-Ping David: “Compact and high resolution virtual mouse using lens array and light sensor”. In: *Three-Dimensional Imaging, Visualization, and Display 2016.* Ed. by Javidi, Bahram and Son, Jung-Young. Vol. 9867. International Society for Optics and Photonics. SPIE, 2016, pp. 322–330. doi: 10.1117/12.2229446. url: https://doi.org/10.1117/12.2229446 (cit. on p. 19).

[Raz01] Razzaque, Sharif; Kohn, Zachariah and Whitton, Mary C: “Redirected Walking”. In: *Proceedings of Eurographics.* Vol. 9. Citeseer. 2001, pp. 105–106 (cit. on pp. 31, 115).

[Reg01] Regenbrecht, Holger; Baratoff, Gregory and Wagner, Michael: “A tangible AR desktop environment”. In: *Computers and Graphics* 25.5 (2001). Mixed realities - beyond conventions, pp. 755–763. doi: https://doi.org/10.1016/S0097-8493(01)00118-2 (cit. on p. 20).

[Reg02a] Regenbrecht, H.T.; Wagner, T. and Baratoff, G.: “MagicMeeting: A Collaborative Tangible Augmented Reality System”. Undetermined. In: *Virtual Reality* 6.3 (2002), pp. 151–166. doi: 10.1007/s100550200016 (cit. on p. 20).

[Reg02b] Regenbrecht, Holger T. and Wagner, Michael T.: “Interaction in a Collaborative Augmented Reality Environment”. In: *CHI ’02 Extended Abstracts on Human Factors in Computing Systems.* CHI EA ’02. Minneapolis, Minnesota, USA: Association for Computing Machinery, 2002, pp. 504–505. doi: 10.1145/506443.506451. url: https://doi.org/10.1145/506443.506451 (cit. on p. 20).

[Reh16] Rehder, J.; Nikolic, J.; Schneider, T.; Hinzmann, T. and Siegwart, R.: “Extending Kalibr: Calibrating the Extrinsics of Multiple IMUs and of Individual Axes”. In: *2016 IEEE International Conference on Robotics and Automation (ICRA).* 2016 IEEE International Conference on Robotics and Automation (ICRA). May 2016, pp. 4304–4311. doi: 10.1109/ICRA.2016.7487628 (cit. on p. 86).

[Rei20] Reinhardt, Jens and Wolf, Katrin: “Go-Through: Disabling Collision to Access Obstructed Paths and Open Occluded Views in Social VR”. In: *Proceedings of the Augmented Humans International Conference.* AHs ’20. Kaiserslautern, Germany: Association for Computing Machinery, 2020. doi: 10.1145/3384657.3384784. url: https://doi.org/10.1145/3384657.3384784 (cit. on p. 24).

[Ren13] Ren, Gang and O’Neill, Eamonn: “Freehand Gestural Text Entry for Interactive TV”. In: *Proceedings of the 11th European Conference on Interactive TV and Video.* EuroITV ’13. Como, Italy: Association for Computing Machinery, 2013, pp. 121–130. doi: 10.1145/2465958.2465966. url: https://doi.org/10.1145/2465958.2465966 (cit. on p. 17).

[Rho16] Rhodin, Helge; Richardt, Christian; Casas, Dan; Insafutdinov, Eldar; Shafiei, Mohammad; Seidel, Hans-Peter; Schiele, Bernt and Theobalt, Christian: “EgoCap: Egocentric Marker-Less Motion Capture with Two Fisheye Cameras”. In: *ACM Trans. Graph.* 35.6 (Nov. 2016). doi: 10.1145/2980179.2980235. url: https://doi.org/10.1145/2980179.2980235 (cit. on pp. 103, 114).

[Rhu14] Rhudy, Matthew: “Time Alignment Techniques for Experimental Sensor Data”. In: *International Journal of Computer Science & Engineering Survey* 5 (Apr. 2014). doi: 10.5121/ijcses.2014.5201 (cit. on p. 82).

[Ria15] Riaz, Qaiser; Tao, Guanhong; Krüger, Björn and Weber, Andreas: “Motion reconstruction using very few accelerometers and ground contacts". In: *Graphical Models* 79 (2015), pp. 23–38. DOI: https://doi.org/10.1016/j.gmod.2015.04.001. URL: http://www.sciencedirect.com/science/article/pii/S1524070315000107 (cit. on p. 103).

[Rie08] Ries, Brian; Interrante, Victoria; Kaeding, Michael and Anderson, Lee: "The Effect of Self-Embodiment on Distance Perception in Immersive Virtual Environments". In: *Proceedings of the 2008 ACM Symposium on Virtual Reality Software and Technology*. VRST '08. Bordeaux, France: Association for Computing Machinery, 2008, pp. 167–170. DOI: 10.1145/1450579.1450614. URL: https://doi.org/10.1145/1450579.1450614 (cit. on p. 13).

[Rie17] Rietzler, Michael; Geiselhart, Florian and Rukzio, Enrico: "The Matrix Has You: Realizing Slow Motion in Full-Body Virtual Reality". In: *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*. ACM. 2017, p. 2 (cit. on p. 31).

[Rie18] Rietzler, Michael; Geiselhart, Florian; Frommel, Julian and Rukzio, Enrico: "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*. ACM. 2018, p. 460 (cit. on p. 31).

[Rin10] Rinck, Mike; Rörtgen, Tobias; Lange, Wolf-Gero; Dotsch, Ron; Wigboldus, Daniël HJ and Becker, Eni S: "Social anxiety predicts avoidance behaviour in virtual encounters". In: *Cognition and Emotion* 24.7 (2010), pp. 1269–1276 (cit. on p. 30).

[Rio18] Rios, Alejandro; Palomar, Marc and Pelechano, Nuria: "Users' Locomotor Behavior in Collaborative Virtual Reality". In: *Proceedings of the 11th Annual International Conference on Motion, Interaction, and Games*. MIG '18. Limassol, Cyprus: Association for Computing Machinery, 2018. DOI: 10.1145/3274247.3274513. URL: https://doi.org/10.1145/3274247.3274513 (cit. on pp. 23, 25).

[Rix16] Rix, Joachim; Haas, Stefan and Teixeira, José: Virtual prototyping: Virtual environments and the product design process. Springer, 2016 (cit. on p. 2).

[Rob00] Robertson, George; Van Dantzich, Maarten; Robbins, Daniel; Czerwinski, Mary; Hinckley, Ken; Risden, Kirsten; Thiel, David and Gorokhovsky, Vadim: "The Task Gallery: a 3D window manager". In: *Proceedings of the SIGCHI conference on Human Factors in Computing Systems.* ACM. 2000, pp. 494–501 (cit. on p. 12).

[Rob03] Roberts, David; Wolff, Robin; Otto, Oliver and Steed, Anthony: "Constructing a Gazebo: Supporting Teamwork in a Tightly Coupled, Distributed Task in Virtual Reality". In: *Presence: Teleoperators & Virtual Environments* 12.6 (2003), pp. 644–657 (cit. on p. 138).

[Rob06] Robler, P. and Hanebeck, U. D.: "Simultaneous Motion Compression for Multi-User Extended Range Telepresence". In: *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems.* 2006, pp. 5189–5194. doi: 10.1109/IROS.2006.281656 (cit. on p. 25).

[Rob16] Robertson, Judy and Kaptein, Maurits: Modern statistical methods for HCI. Springer International Publishing, 2016, pp. 1–348. doi: 10.1007/978-3-319-26633-6. url: https://doi.org/10.1007/978-3-319-26633-6 (cit. on pp. 294, 297).

[Ros99] Rosenberg, Robert and Slater, Mel: "The chording glove: a glove-based text input device". In: *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 29.2 (1999), pp. 186–191 (cit. on pp. 14, 39, 40).

[Rot15] Roth, Daniel; Latoschik, Marc Erich; Vogeley, Kai and Bente, Gary: "Hybrid Avatar-Agent Technology – A Conceptual Step Towards Mediated "Social" Virtual Reality and its Respective Challenges". In: *i-com* 14.2 (Aug. 2015), pp. 107–114. doi: https://doi.org/10.1515/icom-2015-0030. url: https://www.degruyter.com/view/journals/icom/14/2/article-p107.xml (cit. on p. 31).

[Rot16] Roth, D.; Lugrin, J. L.; Galakhov, D.; Hofmann, A.; Bente, G.; Latoschik, M. E. and Fuhrmann, A.: “Avatar realism and social interaction quality in virtual reality”. In: *2016 IEEE Virtual Reality (VR)*. Mar. 2016, pp. 277–278. doi: 10.1109/VR.2016.7504761 (cit. on pp. 104, 114).

[Rot18] Roth, Daniel; Mal, David; Purps, Christian Felix; Kullmann, Peter and Latoschik, Marc Erich: “Injecting Nonverbal Mimicry with Hybrid Avatar-Agent Technologies: A Naïve Approach”. In: *Proceedings of the Symposium on Spatial User Interaction*. ACM. 2018, pp. 69–73 (cit. on pp. 31, 115, 130, 152, 161).

[Rus97] Ruspini, D. C.; Kolarov, K. and Khatib, O.: “Haptic interaction in virtual environments”. In: *Proceedings of the 1997 IEEE/RSJ International Conference on Intelligent Robot and Systems. Innovative Robotics for Real-World Applications. IROS ’97*. Vol. 1. 1997, 128–133 vol.1. doi: 10.1109/IROS.1997.649024 (cit. on pp. 16, 36).

[Sah14] Sahami Shirazi, Alireza; Abdelrahman, Yomna; Henze, Niels; Schneegass, Stefan; Khalilbeigi, Mohammadreza and Schmidt, Albrecht: “Exploiting Thermal Reflection for Interactive Systems”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’14. Toronto, Ontario, Canada: Association for Computing Machinery, 2014, pp. 3483–3492. doi: 10.1145/2556288.2557208. url: https://doi.org/10.1145/2556288.2557208 (cit. on p. 18).

[Sai07] Saito, Shigeru; Hiyama, Atsushi; Tanikawa, Tomohiro and Hirose, Michitaka: “Indoor Marker-Based Localization Using Coded Seamless Pattern for Interior Decoration”. In: *2007 IEEE Virtual Reality Conference*. 2007 IEEE Virtual Reality Conference. Mar. 2007, pp. 67–74. doi: 10.1109/VR.2007.352465 (cit. on p. 23).

[Sai18] Sait, Mohamed Suhail Mohamed Yousuf; Sargunam, Shyam Prathish; Han, Dustin T and Ragan, Eric D: “Physical Hand Interaction for Controlling Multiple Virtual Objects in Virtual

Reality”. In: *Proceedings of the 3rd International Workshop on Interactive and Spatial Computing*. ACM. 2018, pp. 64–74 (cit. on p. 31).

[Sal08] Salzmann, H. and Froehlich, B.: “The Two-User Seating Buck: Enabling Face-to-Face Discussions of Novel Car Interface Concepts”. In: *2008 IEEE Virtual Reality Conference*. 2008, pp. 75–82. doi: 10.1109/VR.2008.4480754 (cit. on p. 17).

[Sal09] Salzmann, H.; Jacobs, J. and Fröhlich, B.: “Collaborative Interaction in Co-located Two-user Scenarios”. In: *Proceedings of the 15th Joint Virtual Reality Eurographics Conference on Virtual Environments*. JVRC’09. Lyon, France: Eurographics Association, 2009, pp. 85–92. doi: 10.2312/EGVE/JVRC09/085-092. url: http://dx.doi.org/10.2312/EGVE/JVRC09/085-092 (cit. on pp. 15, 23, 138).

[San06] Sandnes, Frode Eika and Aubert, Andre: “Bimanual text entry using game controllers: Relying on users’ spatial familiarity with QWERTY”. In: *Interacting with Computers* 19.2 (2006), pp. 140–150 (cit. on pp. 13, 39).

[San10] Sanchez-Vives, Maria V.; Spanlang, Bernhard; Frisoli, Antonio; Bergamasco, Massimo and Slater, Mel: “Virtual Hand Illusion Induced by Visuomotor Correlations”. In: *PLOS ONE* 5.4 (Apr. 2010), pp. 1–6. doi: 10.1371/journal.pone.0010381. url: https://doi.org/10.1371/journal.pone.0010381 (cit. on p. 13).

[San15] Sanz, F. A.; Olivier, A.; Bruder, G.; Pettré, J. and Lécuyer, A.: “Virtual proxemics: Locomotion in the presence of obstacles in large immersive projection environments”. In: *2015 IEEE Virtual Reality (VR)*. 2015, pp. 75–80. doi: 10.1109/VR.2015.7223327 (cit. on p. 25).

[Sar06] Sarakoglou, I.; Bezdicek, M.; Tsagarakis, N. and Caldwell, D. G.: “Free to Touch: A Portable Tactile Display For 3D Surface Texture Exploration”. In: *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2006, pp. 3587–3592. doi: 10.1109/IROS.2006.281649 (cit. on p. 16).

[Sat12] Sato, Munehiko; Poupyrev, Ivan and Harrison, Chris: "Touché: Enhancing Touch Interaction on Humans, Screens, Liquids, and Everyday Objects". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* CHI '12. Austin, Texas, USA: Association for Computing Machinery, 2012, pp. 483–492. doi: 10.1145/2207676.2207743. url: https://doi.org/10.1145/2207676.2207743 (cit. on p. 19).

[Sav12] Savage, Valkyrie; Zhang, Xiaohan and Hartmann, Björn: "Midas: Fabricating Custom Capacitive Touch Sensors to Prototype Interactive Objects". In: *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology.* UIST '12. Cambridge, Massachusetts, USA: Association for Computing Machinery, 2012, pp. 579–588. doi: 10.1145/2380116.2380189. url: https://doi.org/10.1145/2380116.2380189 (cit. on p. 18).

[Sca17] Scavarelli, Anthony and Teather, Robert J.: "VR Collide! Comparing Collision-Avoidance Methods Between Co-Located Virtual Reality Users". In: *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems.* CHI EA '17. Denver, Colorado, USA: Association for Computing Machinery, 2017, pp. 2915–2921. doi: 10.1145/3027063.3053180. url: https://doi.org/10.1145/3027063.3053180 (cit. on pp. 25, 103).

[Sch02] Schmalstieg, Dieter; Fuhrmann, Anton; Hesina, Gerd; Szalavári, Zsolt; Encarnação, L. Miguel; Gervautz, Michael and Purgathofer, Werner: "The Studierstube Augmented Reality Project". In: *Presence: Teleoperators and Virtual Environments* 11.1 (2002), pp. 33–54. doi: 10.1162/105474602317343640. url: https://doi.org/10.1162/105474602317343640 (cit. on p. 17).

[Sch07] Scheibe, R.; Moehring, M. and Froehlich, B.: "Tactile Feedback at the Finger Tips for Improved Direct Interaction in Immersive Environments". In: *2007 IEEE Symposium on 3D User Interfaces.* 2007. doi: 10.1109/3DUI.2007.340784 (cit. on p. 16).

[Sch10] Schultze, Ulrike: “Embodiment and presence in virtual worlds: a review”. In: *Journal of Information Technology* 25.4 (2010), pp. 434–449. doi: 10.1057/jit.2010.25. url: https://doi.org/10.1057/jit.2010.25 (cit. on p. 32).

[Sch14] Schrepp, Martin; Hinderks, Andreas and Thomaschewski, Jörg: “Applying the User Experience Questionnaire (UEQ) in Different Evaluation Scenarios”. In: *Design, User Experience, and Usability. Theories, Methods, and Tools for Designing the User Experience.* Ed. by Marcus, Aaron. Cham: Springer International Publishing, 2014, pp. 383–392 (cit. on p. 293).

[Sch15] Scheggi, Stefano; Meli, Leonardo; Pacchierotti, Claudio and Prattichizzo, Domenico: “Touch the Virtual Reality: Using the Leap Motion Controller for Hand Tracking and Wearable Tactile Devices for Immersive Haptic Rendering”. In: *ACM SIGGRAPH 2015 Posters.* SIGGRAPH ’15. Los Angeles, California: Association for Computing Machinery, 2015. doi: 10.1145/2787626.2792651 (cit. on p. 16).

[Sch16a] Schönberger, Johannes Lutz and Frahm, Jan-Michael: “Structure-from-Motion Revisited”. In: *Conference on Computer Vision and Pattern Recognition (CVPR).* 2016 (cit. on p. 86).

[Sch16b] Schönberger, Johannes Lutz; Zheng, Enliang; Pollefeys, Marc and Frahm, Jan-Michael: “Pixelwise View Selection for Unstructured Multi-View Stereo”. In: *European Conference on Computer Vision (ECCV).* 2016 (cit. on p. 86).

[Sch16c] Schouten, Alexander P; Hooff, Bart van den and Feldberg, Frans: “Virtual team work: Group decision making in 3D virtual environments”. In: *Communication Research* 43.2 (2016), pp. 180–210 (cit. on pp. 3, 22).

[Sch17] Schrepp, Martin; Hinderks, Andreas and Thomaschewski, Jörg: “Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S)”. In: *IJIMAI* 4.6 (2017), pp. 103–108 (cit. on p. 293).

[Sch19] Schneider, D.; Otte, A.; Gesslein, T.; Gagel, P.; Kuth, B.; Damlakhi, M. S.; Dietz, O.; Ofek, E.; Pahud, M.; Kristensson, P. O.; Müller, J. and Grubert, J.: “ReconViguRation: Reconfiguring Physical Keyboards in Virtual Reality”. In: *IEEE Transactions on Visualization and Computer Graphics* 25.11 (2019), pp. 3190–3201. doi: 10.1109/TVCG.2019.2932239 (cit. on pp. 17, 47).

[Sch99] Schmalstieg, Dieter; Encarnação, L. Miguel and Szalavári, Zsolt: “Using Transparent Props for Interaction with the Virtual Table”. In: *Proceedings of the 1999 Symposium on Interactive 3D Graphics.* I3D ’99. Atlanta, Georgia, USA: Association for Computing Machinery, 1999, pp. 147–153. doi: 10.1145/300523.300542. url: https://doi.org/10.1145/300523.300542 (cit. on pp. 17, 18).

[Sco12] Scott, David W.: “Multivariate Density Estimation and Visualization”. In: *Handbook of Computational Statistics.* Ed. by Gentle, James E.; Härdle, Wolfgang Karl and Mori, Yuichi. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 549–569. doi: 10.1007/978-3-642-21551-3_19 (cit. on p. 294).

[Sey17] Seymour, Michael; Riemer, Kai and Kay, Judy: “Interactive Realistic Digital Avatars-Revisiting the Uncanny Valley”. In: *Proceedings of the 50th Hawaii International Conference on System Sciences.* 2017 (cit. on p. 12).

[Sha65] Shapiro, S. S. and Wilk, M. B.: “An Analysis of Variance Test for Normality (Complete Samples)”. In: *Biometrika* 52.3/4 (1965), pp. 591–611. url: http://www.jstor.org/stable/2333709 (cit. on p. 296).

[Shi09] Shintel, Hadas and Keysar, Boaz: “Less Is More: A Minimalist Account of Joint Action in Communication”. In: *Topics in Cognitive Science* 1.2 (2009), pp. 260–273 (cit. on p. 30).

[Shi19] Shigeyama, Jotaro; Hashimoto, Takeru; Yoshida, Shigeo; Narumi, Takuji; Tanikawa, Tomohiro and Hirose, Michitaka: “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. Glasgow, Scotland Uk: Association for Computing Machinery, 2019, pp. 1–11. DOI: 10.1145/3290605.3300241. URL: https://doi.org/10.1145/3290605.3300241 (cit. on p. 13).

[Shi20] SHI, Yilei; ZHANG, Haimo; ZHAO, Kaixing; CAO, Jiashuo; SUN, Mengmeng and NANAYAKKARA, Suranga: "Ready, Steady, Touch! Sensing Physical Contact with a Finger-Mounted IMU". In: *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4.2 (June 2020). DOI: 10.1145/3397309. URL: https://doi.org/10.1145/3397309 (cit. on p. 19).

[Shi93] SHIMOGA, Karun B: "A survey of perceptual feedback issues in dexterous telemanipulation. II. Finger touch feedback". In: *Virtual Reality Annual International Symposium, 1993., 1993 IEEE.* IEEE. 1993, pp. 271–279 (cit. on p. 10).

[Sim15] SIMEONE, Adalberto L.; VELLOSO, Eduardo and GELLERSEN, Hans: "Substitutional Reality: Using the Physical Environment to Design Virtual Reality Experiences". In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems.* CHI '15. Seoul, Republic of Korea: Association for Computing Machinery, 2015, pp. 3307–3316. DOI: 10.1145/2702123.2702389 (cit. on p. 16).

[Sin13] SINCLAIR, Mike; PAHUD, Michel and BENKO, Hrvoje: "TouchMover: Actuated 3D Touchscreen with Haptic Feedback". In: *Proceedings of the 2013 ACM International Conference on Interactive Tabletops and Surfaces.* ITS '13. St. Andrews, Scotland, United Kingdom: Association for Computing Machinery, 2013, pp. 287–296. DOI: 10.1145/2512349.2512805. URL: https://doi.org/10.1145/2512349.2512805 (cit. on pp. 15, 18).

[Sla02] SLATER, Mel and STEED, Anthony: "Meeting People Virtually: Experiments in Shared Virtual Environments". In: *Schroeder R. (eds) The Social Life of Avatars. Presence and Interaction in Shared*

Virtual Environments. Computer Supported Cooperative Work. Springer, 2002, pp. 146–171 (cit. on p. 32).

[Sla16] Slater, Mel and Sanchez-Vives, Maria V: "Enhancing our lives with immersive virtual reality". In: *Frontiers in Robotics and AI* 3 (2016), p. 74 (cit. on pp. 5, 21).

[Sla95] Slater, Mel; Usoh, Martin and Steed, Anthony: "Taking Steps: The Influence of a Walking Technique on Presence in Virtual Reality". In: *ACM Trans. Comput.-Hum. Interact.* 2.3 (Sept. 1995), pp. 201–219. doi: 10.1145/210079.210084. url: https://doi.org/10.1145/210079.210084 (cit. on p. 13).

[Smi18] Smith, Harrison Jesse and Neff, Michael: "Communication Behavior in Embodied Virtual Reality". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.* ACM. 2018, p. 289 (cit. on pp. 13, 32, 163).

[Sod13] Sodhi, Rajinder S; Jones, Brett R; Forsyth, David; Bailey, Brian P and Maciocci, Giuliano: "BeThere: 3D Mobile Collaboration with Spatial Input". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM. 2013, pp. 179–188 (cit. on p. 27).

[Soh17] Sohre, N.; Mackin, C.; Interrante, V. and Guy, S. J.: "Evaluating collision avoidance effects on discomfort in virtual environments". In: *2017 IEEE Virtual Humans and Crowds for Immersive Environments (VHCIE).* 2017, pp. 1–5. doi: 10.1109/VHCIE.2017.7935624 (cit. on p. 24).

[Sol10] Solazzi, M.; Frisoli, A. and Bergamasco, M.: "Design of a novel finger haptic interface for contact and orientation display". In: *2010 IEEE Haptics Symposium.* 2010, pp. 129–132. doi: 10.1109/HAPTIC.2010.5444667 (cit. on p. 16).

[Som59] Sommer, Robert: "Studies in Personal Space". In: *Sociometry* 22.3 (1959), pp. 247–260 (cit. on pp. 30, 137).

[Son11] Song, Peng; Goh, Wooi Boon; Fu, Chi-Wing; Meng, Qiang and Heng, Pheng-Ann: “WYSIWYF: Exploring and Annotating Volume Data with a Tangible Handheld Device”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* CHI ’11. Vancouver, BC, Canada: Association for Computing Machinery, 2011, pp. 1333–1342. doi: 10.1145/1978942.1979140. url: https://doi.org/10.1145/1978942.1979140 (cit. on p. 15).

[Sou01] Soukoreff, R William and MacKenzie, I Scott: “Measuring errors in text entry tasks: an application of the Levenshtein string distance statistic”. In: *CHI’01 extended abstracts on Human factors in computing systems.* ACM. 2001, pp. 319–320 (cit. on pp. 38, 40).

[Sou03] Soukoreff, R William and MacKenzie, I Scott: “Metrics for text entry research: an evaluation of MSD and KSPC, and a new unified error metric”. In: *Proceedings of the SIGCHI conference on Human factors in computing systems.* ACM. 2003, pp. 113–120 (cit. on pp. 38, 40).

[Sou19] Sousa, Maurício; Mendes, Daniel; Anjos, Rafael K. dos; Lopes, Daniel Simões and Jorge, Joaquim: “Negative Space: Workspace Awareness in 3D Face-to-Face Remote Collaboration”. In: *The 17th International Conference on Virtual-Reality Continuum and Its Applications in Industry.* VRCAI ’19. Brisbane, QLD, Australia: Association for Computing Machinery, 2019. doi: 10.1145/3359997.3365744. url: https://doi.org/10.1145/3359997.3365744 (cit. on pp. 32, 115, 148).

[Sou20] Souchet, Alexis; Cassagnac, R; Drapier, O; Maurice, É; Azouani, R; Zaza, E; Naudot, N; Charron, V; Risetti, A; Hono, J and Philippe, Stéphanie: “Virtual Classroom’s Quality of Experience: a collaborative VR platform tested in situ”. In: Apr. 2020 (cit. on pp. 23, 112).

[Spe18] Speicher, Marco; Feit, Anna Maria; Ziegler, Pascal and Krüger, Antonio: “Selection-Based Text Entry in Virtual Reality”. In: *Proceedings of the 2018 CHI Conference on Human*

Factors in Computing Systems. CHI '18. Montreal QC, Canada: Association for Computing Machinery, 2018, pp. 1–13. doi: 10.1145/3173574.3174221. url: https://doi.org/10.1145/3173574.3174221 (cit. on p. 17).

[Spi09] Spindler, Martin; Stellmach, Sophie and Dachselt, Raimund: "PaperLens: Advanced Magic Lens Interaction above the Tabletop". In: *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces.* ITS '09. Banff, Alberta, Canada: Association for Computing Machinery, 2009, pp. 69–76. doi: 10.1145/1731903.1731920. url: https://doi.org/10.1145/1731903.1731920 (cit. on pp. 16, 17).

[Spi10] Spindler, Martin; Tominski, Christian; Schumann, Heidrun and Dachselt, Raimund: "Tangible Views for Information Visualization". In: *ACM International Conference on Interactive Tabletops and Surfaces.* ITS '10. Saarbrücken, Germany: Association for Computing Machinery, 2010, pp. 157–166. doi: 10.1145/1936652.1936684. url: https://doi.org/10.1145/1936652.1936684 (cit. on p. 15).

[Spi14] Spindler, Martin; Büschel, Wolfgang; Winkler, Charlotte and Dachselt, Raimund: "Tangible Displays for the Masses: Spatial Interaction with Handheld Displays by Using Consumer Depth Cameras". In: *Personal Ubiquitous Comput.* 18.5 (June 2014), pp. 1213–1225. doi: 10.1007/s00779-013-0730-7. url: https://doi.org/10.1007/s00779-013-0730-7 (cit. on p. 11).

[Sri17] Sridhar, Srinath; Markussen, Anders; Oulasvirta, Antti; Theobalt, Christian and Boring, Sebastian: "WatchSense: On- and Above-Skin Input Sensing through a Wearable Depth Sensor". In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems.* CHI '17. Denver, Colorado, USA: Association for Computing Machinery, 2017, pp. 3891–3902. doi: 10.1145/3025453.3026005. url: https://doi.org/10.1145/3025453.3026005 (cit. on p. 19).

[Sta19] Starke, Sebastian; Zhang, He; Komura, Taku and Saito, Jun: “Neural State Machine for Character-Scene Interactions”. In: *ACM Trans. Graph.* 38.6 (Nov. 2019). doi: 10.1145/3355089.3356505. url: https://doi.org/10.1145/3355089.3356505 (cit. on p. 179).

[Ste03] Steed, Anthony; Spante, Maria; Heldal, Ilona; Axelsson, Ann-Sofie and Schroeder, Ralph: “Strangers and Friends in Caves: An Exploratory Study of Collaboration in Networked IPT Systems for Extended Periods of Time”. In: *Proceedings of the 2003 Symposium on Interactive 3D Graphics.* ACM. 2003, pp. 51–54 (cit. on p. 138).

[Ste05] Steinicke, Frank; Ropinski, Timo and Hinrichs, Klaus: “VR and laser-based interaction in virtual environments using a dual-purpose interaction metaphor”. In: *IEEE VR 2005 Workshop Proceedings on New Directions in 3D User Interfaces.* 2005, pp. 61–64 (cit. on p. 20).

[Ste13] Steimle, Jürgen; Jordt, Andreas and Maes, Pattie: “Flexpad: Highly Flexible Bending Interactions for Projected Handheld Displays”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* CHI ’13. Paris, France: Association for Computing Machinery, 2013, pp. 237–246. doi: 10.1145/2470654.2470688. url: https://doi.org/10.1145/2470654.2470688 (cit. on p. 17).

[Ste15] Steed, Anthony and Schroeder, Ralph: “Collaboration in Immersive and Non-immersive Virtual Environments”. In: *Immersed in Media.* Springer, 2015, pp. 263–282 (cit. on p. 104).

[Ste16] Steed, A.; Pan, Y.; Zisch, F. and Steptoe, W.: “The impact of a self-avatar on cognitive load in immersive virtual reality”. In: *2016 IEEE Virtual Reality (VR).* 2016, pp. 67–76. doi: 10.1109/VR.2016.7504689 (cit. on p. 13).

[Ste92] Steuer, Jonathan: “Defining virtual reality: Dimensions determining telepresence”. In: *Journal of communication* 42.4 (1992), pp. 73–93 (cit. on p. 9).

[Sti02] Stiefelhagen, Rainer and Zhu, Jie: “Head Orientation and Gaze Direction in Meetings”. In: *CHI’02 Extended Abstracts on Human Factors in Computing Systems.* ACM. 2002, pp. 858–859 (cit. on pp. 141, 170).

[Sto70] Stoudt, Howard W: “Skinfolds, Body Girths, Biacromial Diameter, and Selected Anthropometric Indices of Adults, United States, 1960-1962”. In: *Vital and health statistics. Series 11, Data from the national health survey; no. 35; Public Health Service publication; no. 1000, ser. 11, no. 35;* (1970) (cit. on p. 141).

[Sto95] Stoakley, Richard; Conway, Matthew J and Pausch, Randy: “Virtual reality on a WIM: interactive worlds in miniature”. In: *Proceedings of the SIGCHI conference on Human factors in computing systems.* ACM Press/Addison-Wesley Publishing Co. 1995, pp. 265–272 (cit. on pp. 15, 31).

[Str09] Streuber, Stephan; Rosa, Stephan de la; Trutoiu, Laura; Bülthoff, Heinrich H. and Mohler, Betty J.: “Does Brief Exposure to a Self-avatar Effect Common Human Behaviors in Immersive Virtual Environments?” In: *Eurographics 2009 - Short Papers.* Ed. by Alliez, P. and Magnor, M. European Association for Computer Graphics. Geneve, Switzerland: The Eurographics Association, 2009, pp. 33–36. doi: 10.2312/egs.20091042. url: http://dx.doi.org/10.2312/egs.20091042 (cit. on p. 13).

[Stu94a] Sturman, D. J. and Zeltzer, D.: “A survey of glove-based input”. In: *IEEE Computer Graphics and Applications* 14.1 (1994), pp. 30–39. doi: 10.1109/38.250916 (cit. on p. 1).

[Stu94b] Sturman, David J and Zeltzer, David: “A survey of glove-based input”. In: *Computer Graphics and Applications, IEEE* 14.1 (1994), pp. 30–39 (cit. on p. 11).

[Su07] Su, Jianbo: “Motion Compression for Telepresence Locomotion”. In: *Presence: Teleoperators and Virtual Environments* 16.4 (2007), pp. 385–398. doi: 10.1162/pres.16.4.385. url: https://doi.org/10.1162/pres.16.4.385 (cit. on p. 25).

[Sur19] Surale, Hemant Bhaskar; Gupta, Aakar; Hancock, Mark and Vogel, Daniel: "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. Glasgow, Scotland Uk: Association for Computing Machinery, 2019, pp. 1–13. doi: 10.1145/3290605.3300243. url: https://doi.org/10.1145/3290605.3300243 (cit. on pp. 15, 18).

[Sut65] Sutherland, Ivan E: "The ultimate display". In: *Multimedia: From Wagner to virtual reality* 1 (1965) (cit. on p. 1).

[Sut68] Sutherland, Ivan E.: "A Head-Mounted Three Dimensional Display". In: *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I.* AFIPS '68 (Fall, part I). San Francisco, California: Association for Computing Machinery, 1968, pp. 757–764. doi: 10.1145/1476589.1476686. url: https://doi.org/10.1145/1476589.1476686 (cit. on p. 1).

[Syk20] Sykownik, Philipp; Emmerich, Katharina and Masuch, Maic: "Like in the Good Old Times, but Virtual - A Case for Simulating Co-Located Multiplayer Games in VR". In: *Extended Abstracts of the 2020 Annual Symposium on Computer-Human Interaction in Play*. New York, NY, USA: Association for Computing Machinery, 2020, pp. 379–383. url: https://doi.org/10.1145/3383668.3419885 (cit. on pp. 22, 112).

[Sza97] Szalavári, Zsolt and Gervautz, Michael: "The Personal Interaction Panel – a Two-Handed Interface for Augmented Reality". In: *Computer Graphics Forum* 16.3 (1997), pp. C335–C346. doi: https://doi.org/10.1111/1467-8659.00137. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-8659.00137. url: https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-8659.00137 (cit. on p. 18).

[Sza98] Szalavári, Zsolt; Schmalstieg, Dieter; Fuhrmann, Anton and Gervautz, Michael: ""Studierstube": An environment for collaboration in augmented reality". In: *Virtual Reality* 3.1 (1998), pp. 37–48 (cit. on p. 21).

[Tak15] Takanaka, Shun; Yano, Hiroaki and Iwata, Hiroo: “Multitouch Haptic Interface with Movable Touch Screen”. In: *SIGGRAPH Asia 2015 Haptic Media And Contents Design.* SA ’15. Kobe, Japan: Association for Computing Machinery, 2015. doi: 10.1145/2818384.2818396. url: https://doi.org/10.1145/2818384.2818396 (cit. on pp. 15, 18).

[Tak17] Taketomi, Takafumi; Uchiyama, Hideaki and Ikeda, Sei: “Visual SLAM Algorithms: A Survey from 2010 to 2016”. In: *IPSJ Transactions on Computer Vision and Applications* 9.1 (Dec. 2017), p. 16. doi: 10.1186/s41074-017-0027-2 (cit. on p. 24).

[Tak18] Takanaka, Shun; Yano, Hiroaki and Iwata, Hiroo: “3DOF Multitouch Haptic Interface with Movable Touchscreen”. In: *Haptic Interaction.* Ed. by Hasegawa, Shoichi; Konyo, Masashi; Kyung, Ki-Uk; Nojima, Takuya and Kajimoto, Hiroyuki. Singapore: Springer Singapore, 2018, pp. 55–61 (cit. on pp. 15, 18).

[Tan10] Tang, Anthony; Pahud, Michel; Inkpen, Kori; Benko, Hrvoje; Tang, John C. and Buxton, Bill: “Three’s Company: Understanding Communication Channels in Three-Way Distributed Collaboration”. In: *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work.* CSCW ’10. Savannah, Georgia, USA: Association for Computing Machinery, 2010, pp. 271–280. doi: 10.1145/1718918.1718969 (cit. on pp. 30, 32, 137, 148, 149).

[Tan11] Tang, H.H.; Lee, Y.Y. and Gero, J.S.: “Comparing collaborative co-located and distributed design processes in digital and traditional sketching environments: A protocol study using the function–behaviour–structure coding scheme”. In: *Design Studies* 32.1 (2011), pp. 1–29. doi: https://doi.org/10.1016/j.destud.2010.06.004. url: http://www.sciencedirect.com/science/article/pii/S0142694X10000475 (cit. on pp. 22, 112).

[Tan91] Tang, John C and Minneman, Scott: “VideoWhiteboard: video shadows to support remote collaboration”. In: *Proceedings of* *the SIGCHI Conference on Human Factors in Computing Systems.* ACM. 1991, pp. 315–322 (cit. on p. 26).

[Tec12] Tecchia, Franco; Alem, Leila and Huang, Weidong: “3D Helping Hands: A Gesture based MR System for Remote Collaboration”. In: *Proceedings of the 11th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry.* ACM. 2012, pp. 323–328 (cit. on p. 28).

[Teo18] Teo, Theophilus; Lee, Gun A; Billinghurst, Mark and Adcock, Matt: “Hand Gestures and Visual Annotation in Live 360 Panorama-Based Mixed Reality Remote Collaboration”. In: *Proceedings of the 30th Australian Conference on Computer-Human Interaction.* ACM. 2018, pp. 406–410 (cit. on pp. 27, 138).

[Tho19] Thoravi Kumaravel, Balasaravanan; Anderson, Fraser; Fitzmaurice, George; Hartmann, Bjoern and Grossman, Tovi: “Loki: Facilitating Remote Instruction of Physical Tasks Using Bi-Directional Mixed-Reality Telepresence”. In: *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology.* UIST ’19. New Orleans, LA, USA: Association for Computing Machinery, 2019, pp. 161–174. doi: 10.1145/3332165.3347872 (cit. on p. 27).

[Tom14] Tominski, Christian; Gladisch, Stefan; Kister, Ulrike; Dachselt, Raimund and Schumann, Heidrun: “A Survey on Interactive Lenses in Visualization”. In: *EuroVis - STARs.* Ed. by Borgo, R.; Maciejewski, R. and Viola, I. The Eurographics Association, 2014. doi: 10.2312/eurovisstar.20141172 (cit. on p. 16).

[Tom17] Tominski, C.; Gladisch, S.; Kister, U.; Dachselt, R. and Schumann, H.: “Interactive Lenses for Visualization: An Extended Survey”. In: *Computer Graphics Forum* 36.6 (2017), pp. 173–200. doi: https://doi.org/10.1111/cgf.12871. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.12871. url: https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.12871 (cit. on p. 16).

[Toy18] Toyama, Shuma; Al Sada, Mohammed and Nakajima, Tatsuo: "VRowser: A Virtual Reality Parallel Web Browser". In: *Virtual, Augmented and Mixed Reality: Interaction, Navigation, Visualization, Embodiment, and Simulation.* Ed. by Chen, Jessie Y.C. and Fragomeni, Gino. Cham: Springer International Publishing, 2018, pp. 230–244 (cit. on pp. 20, 21).

[Tre06] Tredinnick, Ross; Anderson, Lee; Ries, Brian; Interrante, Victoria and Colucci, D'nardo: "A tablet based immersive architectural design tool". English (US). In: *Synthetic Landscapes - ACADIA 2006 International Conference.* Synthetic Landscapes - ACADIA 2006 International Conference. Synthetic Landscapes - ACADIA 2006 International Conference ; Conference date: 12-10-2006 Through 15-10-2006. Dec. 2006, pp. 328–341 (cit. on p. 19).

[Tse14] Tsetserukou, D.; Hosokawa, S. and Terashima, K.: "LinkTouch: A wearable haptic device with five-bar linkage mechanism for presentation of two-DOF force feedback at the fingerpad". In: *2014 IEEE Haptics Symposium (HAPTICS).* 2014, pp. 307–312. doi: 10.1109/HAPTICS.2014.6775473 (cit. on p. 16).

[Tuk53] Tukey, JW: The Problem of Multiple Comparisons, Princeton University, Princeton, NJ Cited in: Principles and Procedures of Statistics. RG Steele and JH Torrie. 1953 (cit. on p. 296).

[Tza08] Tzafestas, C. S.; Birbas, K.; Koumpouros, Y. and Christopoulos, D.: "Pilot Evaluation Study of a Virtual Paracentesis Simulator for Skill Training and Assessment: The Beneficial Effect of Haptic Display". In: *Presence* 17.2 (Apr. 2008), pp. 212–229. doi: 10.1162/pres.17.2.212 (cit. on p. 15).

[Ull09] Ullrich, Torsten; Settgast, Volker; Ofenböck, Christian and Fellner, Dieter W.: "Desktop Integration in Graphics Environments". In: *Proceedings of the 15th Joint Virtual Reality Eurographics Conference on Virtual Environments.* JVRC'09. Lyon, France: Eurographics Association, 2009, pp. 109–112 (cit. on p. 21).

[Ull79] Ullman, Shimon: “The interpretation of structure from motion”. In: *Proceedings of the Royal Society of London. Series B. Biological Sciences* 203.1153 (1979), pp. 405–426 (cit. on p. 86).

[Ull97] Ullmer, Brygg and Ishii, Hiroshi: “The MetaDESK: Models and Prototypes for Tangible User Interfaces”. In: *Proceedings of the 10th Annual ACM Symposium on User Interface Software and Technology.* UIST ’97. Banff, Alberta, Canada: Association for Computing Machinery, 1997, pp. 223–232. doi: 10.1145/263407.263551. url: https://doi.org/10.1145/263407.263551 (cit. on p. 18).

[Uso99] Usoh, Martin; Arthur, Kevin; Whitton, Mary C.; Bastos, Rui; Steed, Anthony; Slater, Mel and Brooks, Frederick P.: “Walking > Walking-in-Place > Flying, in Virtual Environments”. In: *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques.* SIGGRAPH ’99. USA: ACM Press/Addison-Wesley Publishing Co., 1999, pp. 359–364. doi: 10.1145/311535.311589. url: https://doi.org/10.1145/311535.311589 (cit. on p. 13).

[van13] van de Camp, F. and Stiefelhagen, R.: “Applying force fields to black-box guls using computer vision”. In: *2013 1st IEEE Workshop on User-Centered Computer Vision (UCCV).* Jan. 2013, pp. 1–6. doi: 10.1109/UCCV.2013.6530799 (cit. on p. 20).

[van20] van de Camp, Florian; Hoppe, Adrian H.; Unmüßig, Gabriel and Peinsipp-Byma, Elisabeth: “Cooperative and location-independent terrain assessment for deployment planning using a 3D mixed reality environment”. In: *Virtual, Augmented, and Mixed Reality (XR) Technology for Multi-Domain Operations.* Vol. 11426. International Society for Optics and Photonics. SPIE, 2020, pp. 1–8. doi: 10.1117/12.2558593 (cit. on pp. 10, 165).

[Ven12] Venkatesh, Viswanath and Windeler, Jamie B: “Hype or help? A longitudinal field study of virtual world use for team collaboration”. In: *Journal of the Association for Information Systems* 13.10 (2012), p. 5 (cit. on p. 3).

[Ver02] Vertegaal, Roel and Ding, Yaping: “Explaining Effects of Eye Gaze on Mediated Group Conversations: Amount or Synchronization?” In: *Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work*. ACM. 2002, pp. 41–48 (cit. on p. 31).

[Ver99] Vertegaal, Roel: “The GAZE groupware system: mediating joint attention in multiparty communication and collaboration”. In: *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM. 1999, pp. 294–301 (cit. on p. 27).

[Vin16] Vinayak; Ramanujan, Devarajan; Piya, Cecil and Ramani, Karthik: “MobiSweep: Exploring Spatial Design Ideation Using a Smartphone as a Hand-Held Reference Plane”. In: *Proceedings of the TEI ’16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction*. TEI ’16. Eindhoven, Netherlands: Association for Computing Machinery, 2016, pp. 12–20. doi: 10.1145/2839462.2839490. url: https://doi.org/10.1145/2839462.2839490 (cit. on p. 15).

[Wah16] Wahlström, J.; Skog, I.; Händel, P. and Nehorai, A.: “IMU-Based Smartphone-to-Vehicle Positioning”. In: *IEEE Transactions on Intelligent Vehicles* 1.2 (June 2016), pp. 139–147. doi: 10.1109/TIV.2016.2588978 (cit. on p. 24).

[Wal09] Walker, Guy H.; Stanton, Neville A.; Salmon, Paul and Jenkins, Dan: “How can we support the commander’s involvement in the planning process? An exploratory study into remote and co-located command planning”. In: *International Journal of Industrial Ergonomics* 39.2 (2009), pp. 456–464. doi: https://doi.org/10.1016/j.ergon.2008.12.003. url: http://www.sciencedirect.com/science/article/pii/S0169814108001935 (cit. on pp. 22, 112).

[Wal17] Walker, James; Li, Bochao; Vertanen, Keith and Kuhl, Scott: "Efficient Typing on a Visually Occluded Physical Keyboard". In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems.* ACM. 2017, pp. 5457–5461 (cit. on pp. 13, 39).

[Wan14] Wang, Junjue; Zhao, Kaichen; Zhang, Xinyu and Peng, Chunyi: "Ubiquitous Keyboard for Small Mobile Devices: Harnessing Multipath Fading for Fine-Grained Keystroke Localization". In: *Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services.* MobiSys '14. Bretton Woods, New Hampshire, USA: Association for Computing Machinery, 2014, pp. 14–27. doi: 10.1145/2594368.2594384. url: https://doi.org/10.1145/2594368.2594384 (cit. on p. 17).

[Wan15a] Wang, J. and Lindeman, R. W.: "Object impersonation: Towards effective interaction in tablet- and HMD-based hybrid virtual environments". In: *2015 IEEE Virtual Reality (VR).* 2015, pp. 111–118. doi: 10.1109/VR.2015.7223332 (cit. on p. 19).

[Wan15b] Wang, Jia and Lindeman, Robert: "Coordinated hybrid virtual environments: Seamless interaction contexts for effective virtual reality". In: *Computers & Graphics* 48 (2015), pp. 71–83. doi: https://doi.org/10.1016/j.cag.2015.02.007. url: http://www.sciencedirect.com/science/article/pii/S0097849315000151 (cit. on p. 19).

[Wan16] Wang, John and Olson, Edwin: "AprilTag 2: Efficient and robust fiducial detection". In: *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).* Oct. 2016 (cit. on p. 86).

[Wan19a] Wang, Dangxiao; Guo, Yuan; Liu, Shiyi; Zhang, Yuru; Xu, Weiliang and Xiao, Jing: "Haptic display for virtual reality: progress and challenges". In: *Virtual Reality & Intelligent Hardware* 1.2 (2019), pp. 136–162. doi: https://doi.org/10.3724/SP.J.2096-5796.2019.0008. url: http://www.sciencedirect.com/science/article/pii/S2096579619300130 (cit. on p. 16).

[Wan19b] WANG, Peng; ZHANG, Shusheng; BILLINGHURST, Mark; BAI, Xiaoliang; HE, Weiping; WANG, Shuxia; SUN, Mengmeng and ZHANG, Xu: “A comprehensive survey of AR/MR-based co-design in manufacturing”. In: *Engineering with Computers* (2019), pp. 1–24 (cit. on p. 3).

[Wan20] WANG, L.; WU, Wentao; ZHOU, Zijing and POPESCU, Voicu: “View Splicing for Effective VR Collaboration”. In: 2020 (cit. on p. 30).

[Wat99] WATSEN, Kent; DARKEN, Rudolph and CAPPS, Michael: “A Handheld Computer as an Interaction Device to a Virtual Environment”. In: *Proceedings of the international projection technologies workshop.* Aug. 1999, pp. 303–304 (cit. on p. 19).

[Wei09] WEISS, Malte; WAGNER, Julie; JANSEN, Yvonne; JENNINGS, Roger; KHOSHABEH, Ramsin; HOLLAN, James D. and BORCHERS, Jan: “SLAP Widgets: Bridging the Gap between Virtual and Physical Controls on Tabletops”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* CHI ’09. Boston, MA, USA: Association for Computing Machinery, 2009, pp. 481–490. DOI: 10.1145/1518701.1518779. URL: https://doi.org/10.1145/1518701.1518779 (cit. on p. 47).

[Wei15] WEIGEL, Martin; LU, Tong; BAILLY, Gilles; OULASVIRTA, Antti; MAJIDI, Carmel and STEIMLE, Jürgen: “ISkin: Flexible, Stretchable and Visually Customizable On-Body Touch Sensors for Mobile Computing”. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems.* CHI ’15. Seoul, Republic of Korea: Association for Computing Machinery, 2015, pp. 2991–3000. DOI: 10.1145/2702123.2702391. URL: https://doi.org/10.1145/2702123.2702391 (cit. on pp. 16, 18).

[Wei19a] WEI, Shih-En; SARAGIH, Jason; SIMON, Tomas; HARLEY, Adam W.; LOMBARDI, Stephen; PERDOCH, Michal; HYPES, Alexander; WANG, Dawei; BADINO, Hernan and SHEIKH, Yaser: “VR Facial Animation via Multiview Image Translation”. In: *ACM Trans.*

Graph. 38.4 (July 2019). DOI: 10.1145/3306346.3323030. URL: https://doi.org/10.1145/3306346.3323030 (cit. on p. 114).

[Wei19b] WEISSKER, T.; KULIK, A. and FROEHLICH, B.: "Multi-Ray Jumping: Comprehensible Group Navigation for Collocated Users in Immersive Virtual Reality". In: *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR).* 2019, pp. 136–144. DOI: 10.1109/VR.2019.8797807 (cit. on p. 26).

[Wei20a] WEISSKER, T.; BIMBERG, P. and FROEHLICH, B.: "Getting There Together: Group Navigation in Distributed Virtual Environments". In: *IEEE Transactions on Visualization and Computer Graphics* 26.5 (2020), pp. 1860–1870. DOI: 10.1109/TVCG.2020.2973474 (cit. on p. 26).

[Wei20b] WEISSKER, T.; TORNOW, P. and FROEHLICH, B.: "Tracking Multiple Collocated HTC Vive Setups in a Common Coordinate System". In: *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW).* 2020, pp. 592–593. DOI: 10.1109/VRW50115.2020.00147 (cit. on p. 23).

[Whi18] WHITMIRE, Eric; BENKO, Hrvoje; HOLZ, Christian; OFEK, Eyal and SINCLAIR, Mike: "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. Montreal QC, Canada: Association for Computing Machinery, 2018, pp. 1–12. DOI: 10.1145/3173574.3173660. URL: https://doi.org/10.1145/3173574.3173660 (cit. on p. 16).

[Wid00] WIDESTRÖM, Josef; AXELSSON, Ann-Sofie; SCHROEDER, Ralph; NILSSON, Alexander; HELDAL, Ilona and ABELIN, Åsa: "The Collaborative Cube Puzzle: A Comparison of Virtual and Real Environments". In: *Proceedings of the third International Conference on Collaborative Virtual Environments.* ACM. 2000, pp. 165–171 (cit. on p. 138).

[Wil06] Wilson, Andrew D and Agrawala, Maneesh: "Text entry using a dual joystick game controller". In: *Proceedings of the SIGCHI conference on Human Factors in computing systems.* ACM. 2006, pp. 475–478 (cit. on pp. 13, 39, 40).

[Wil07] Wilson, A. D.: "Depth-Sensing Video Cameras for 3D Tangible Tabletop Interaction". In: *Second Annual IEEE International Workshop on Horizontal Interactive Human-Computer Systems (TABLETOP'07).* 2007, pp. 201–204. doi: 10.1109/TABLETOP.2007.35 (cit. on p. 18).

[Wil10a] Wilson, Andrew D.: "Using a Depth Camera as a Touch Sensor". In: *ACM International Conference on Interactive Tabletops and Surfaces.* ITS '10. Saarbrücken, Germany: Association for Computing Machinery, 2010, pp. 69–72. doi: 10.1145/1936652.1936665. url: https://doi.org/10.1145/1936652.1936665 (cit. on p. 18).

[Wil10b] Wilson, Andrew D. and Benko, Hrvoje: "Combining Multiple Depth Cameras and Projectors for Interactions on, above and between Surfaces". In: *Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology.* UIST '10. New York, New York, USA: Association for Computing Machinery, 2010, pp. 273–282. doi: 10.1145/1866029.1866073. url: https://doi.org/10.1145/1866029.1866073 (cit. on p. 18).

[Wil13] Willis, Karl D. D.; Shiratori, Takaaki and Mahler, Moshe: "HideOut: Mobile Projector Interaction with Tangible Objects and Surfaces". In: *Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction.* TEI '13. Barcelona, Spain: Association for Computing Machinery, 2013, pp. 331–338. doi: 10.1145/2460625.2460682. url: https://doi.org/10.1145/2460625.2460682 (cit. on p. 48).

[Wil19] Willich, Julius von; Funk, Markus; Müller, Florian; Marky, Karola; Riemann, Jan and Mühlhäuser, Max: "You Invaded My Tracking Space! Using Augmented Virtuality for Spotting Passersby in Room-Scale Virtual Reality". In: *Proceedings of the*

2019 on Designing Interactive Systems Conference. DIS '19. San Diego, CA, USA: Association for Computing Machinery, 2019, pp. 487–496. DOI: 10.1145/3322276.3322334. URL: https://doi.org/10.1145/3322276.3322334 (cit. on pp. 25, 103).

[Wil70] WILCOXON, Frank; KATTI, SK and WILCOX, Roberta A: “Critical Values and Probability Levels for the Wilcoxon Rank Sum Test and the Wilcoxon Signed Rank Test”. In: *Selected Tables in Mathematical Statistics* 1 (1970), pp. 171–259 (cit. on p. 296).

[Win02] WINGRAVE, Chadwick A; BOWMAN, Doug A and RAMAKRISHNAN, Naren: “Towards preferences in virtual environment interfaces”. In: *EGVE*. Vol. 2. 2002, pp. 63–72 (cit. on p. 106).

[Win17] WINGERT, Benjamin; SCHÖLLHORN, Isabel and BUES, Matthias: “ProDesk: An Interactive Ubiquitous Desktop Surface”. In: *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces*. ISS '17. Brighton, United Kingdom: Association for Computing Machinery, 2017, pp. 366–371. DOI: 10.1145/3132272.3135078. URL: https://doi.org/10.1145/3132272.3135078 (cit. on p. 18).

[Win40] WINN, Lelia J. and SCHWARZ, Edward R.: “Technical Evaluation of Textile Finishing Treatments: III. Use of Rank Correlation for Comparison of Data”. In: *Textile Research* 10.11 (1940), pp. 453–461. DOI: 10.1177/004051754001001104. URL: https://doi.org/10.1177/004051754001001104 (cit. on p. 298).

[Wlo95a] WLOKA, Matthias M and GREENFIELD, Eliot: “The virtual tricorder: a uniform interface for virtual reality”. In: *Proceedings of the 8th annual ACM symposium on User interface and software technology*. ACM. 1995, pp. 39–40 (cit. on pp. 15, 16).

[Wlo95b] WLOKA, Matthias M. and GREENFIELD, Eliot: “The Virtual Tricorder: A Uniform Interface for Virtual Reality”. In: *Proceedings of the 8th Annual ACM Symposium on User Interface and Software Technology*. UIST '95. Pittsburgh, Pennsylvania, USA: Association for Computing Machinery, 1995, pp. 39–40. DOI: 10.

1145/215585.215647. URL: https://doi.org/10.1145/215585.215647 (cit. on p. 13).

[Wob04] Wobbrock, Jacob O; Myers, Brad A and Aung, Htet Htet: "Writing with a joystick: a comparison of date stamp, selection keyboard, and EdgeWrite". In: *Proceedings of Graphics Interface 2004.* Canadian Human-Computer Communications Society. 2004, pp. 1–8 (cit. on pp. 13, 39, 40).

[Wob09] Wobbrock, Jacob O.; Morris, Meredith Ringel and Wilson, Andrew D.: "User-Defined Gestures for Surface Computing". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* CHI '09. Boston, MA, USA: Association for Computing Machinery, 2009, pp. 1083–1092. DOI: 10.1145/1518701.1518866 (cit. on p. 20).

[Won10] Wong, Nelson and Gutwin, Carl: "Where Are You Pointing?: The Accuracy of Deictic Pointing in CVEs". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM. 2010, pp. 1029–1038 (cit. on p. 132).

[Won14] Wong, Nelson and Gutwin, Carl: "Support for Deictic Pointing in CVEs: Still Fragmented after All These Years?" In: *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing.* ACM. 2014, pp. 1377–1387 (cit. on p. 132).

[Won16] Wong-Villacres, Marisol; Barzola, Vanessa Echeverría; Granda, Roger and Garcia, Katherine Chiluiza: "Portable Tabletops: A Low-Cost Pen-and-Touch Approach". In: *Human-Computer Interaction. Interaction Platforms and Techniques.* Ed. by Kurosu, Masaaki. Cham: Springer International Publishing, 2016, pp. 241–252 (cit. on p. 18).

[Woz18] Wozniak, Peter; Capobianco, Antonio; Javahiraly, Nicolas and Curticapean, Dan: "Towards Unobtrusive Obstacle Detection and Notification for VR". In: *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology.* VRST '18. Tokyo, Japan: Association for Computing Machinery, 2018.

DOI: 10.1145/3281505.3283391. URL: https://doi.org/10.1145/3281505.3283391 (cit. on p. 24).

[Woz20] WOZNIAK, Peter; CAPOBIANCO, Antonio; JAVAHIRALY, Nicolas and CURTICAPEAN, Dan: “Depth Sensor Based Detection of Obstacles and Notification for Virtual Reality Systems”. In: *Advances in Human Factors in Wearable Technologies and Game Design.* Ed. by AHRAM, Tareq. Cham: Springer International Publishing, 2020, pp. 271–282 (cit. on p. 24).

[Wu17] WU, Chien-Min; HSU, Chih-Wen; LEE, Tzu-Kuei and SMITH, Shana: “A Virtual Reality Keyboard with Realistic Haptic Feedback in a Fully Immersive Virtual Environment”. In: *Virtual Real.* 21.1 (Mar. 2017), pp. 19–29. DOI: 10.1007/s10055-016-0296-6. URL: https://doi.org/10.1007/s10055-016-0296-6 (cit. on p. 17).

[Xia13] XIAO, Robert; HARRISON, Chris and HUDSON, Scott E.: “WorldKit: Rapid and Easy Creation of Ad-Hoc Interactive Applications on Everyday Surfaces”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* CHI ’13. Paris, France: Association for Computing Machinery, 2013, pp. 879–888. DOI: 10.1145/2470654.2466113. URL: https://doi.org/10.1145/2470654.2466113 (cit. on p. 17).

[Xia14] XIA, Haijun; JOTA, Ricardo; MCCANNY, Benjamin; YU, Zhe; FORLINES, Clifton; SINGH, Karan and WIGDOR, Daniel: “Zero-Latency Tapping: Using Hover Information to Predict Touch Locations and Eliminate Touchdown Latency”. In: *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology.* UIST ’14. Honolulu, Hawaii, USA: Association for Computing Machinery, 2014, pp. 205–214. DOI: 10.1145/2642918.2647348. URL: https://doi.org/10.1145/2642918.2647348 (cit. on p. 19).

[Xia16] XIAO, Robert; HUDSON, Scott and HARRISON, Chris: “DIRECT: Making Touch Tracking on Ordinary Surfaces Practical with Hybrid Depth-Infrared Sensing”. In: *Proceedings of the 2016*

ACM International Conference on Interactive Surfaces and Spaces. ISS '16. Niagara Falls, Ontario, Canada: Association for Computing Machinery, 2016, pp. 85–94. DOI: 10.1145/2992154.2992173. URL: https://doi.org/10.1145/2992154.2992173 (cit. on p. 17).

[Xia18a] Xia, Haijun; Herscher, Sebastian; Perlin, Ken and Wigdor, Daniel: "Spacetime: Enabling Fluid Individual and Collaborative Editing in Virtual Reality". In: *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology.* UIST '18. Berlin, Germany: Association for Computing Machinery, 2018, pp. 853–866. DOI: 10.1145/3242587.3242597 (cit. on p. 33).

[Xia18b] Xiao, R.; Schwarz, J.; Throm, N.; Wilson, A. D. and Benko, H.: "MRTouch: Adding Touch Input to Head-Mounted Mixed Reality". In: *IEEE Transactions on Visualization and Computer Graphics* 24.4 (Apr. 2018), pp. 1653–1660. DOI: 10.1109/TVCG.2018.2794222 (cit. on pp. 18, 48, 54, 55, 66).

[Xia18c] Xiao, Robert; Cao, Teng; Guo, Ning; Zhuo, Jun; Zhang, Yang and Harrison, Chris: "LumiWatch: On-Arm Projected Graphics and Touch Input". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.* CHI '18. Montreal QC, Canada: Association for Computing Machinery, 2018, pp. 1–11. DOI: 10.1145/3173574.3173669. URL: https://doi.org/10.1145/3173574.3173669 (cit. on p. 16).

[Yam18] Yamamoto, Takuya; Otsuki, Mai; Kuzuoka, Hideaki and Suzuki, Yusuke: "Tele-Guidance System to Support Anticipation during Communication". In: *Multimodal Technologies and Interaction* 2.3 (2018), p. 55 (cit. on p. 32).

[Yan12] Yang, Xing-Dong; Grossman, Tovi; Wigdor, Daniel and Fitzmaurice, George: "Magic Finger: Always-Available Input through Finger Instrumentation". In: *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology.* UIST '12. Cambridge, Massachusetts, USA: Association for Computing Machinery, 2012, pp. 147–156. DOI: 10.1145/2380116.

2380137. URL: https://doi.org/10.1145/2380116.2380137 (cit. on p. 19).

[Yan18] YANG, Jackie Junrui; HORII, Hiroshi; THAYER, Alexander and BALLAGAS, Rafael: "VR Grabbers: Ungrounded Haptic Retargeting for Precision Grabbing Tools". In: *The 31st Annual ACM Symposium on User Interface Software and Technology.* ACM. 2018, pp. 889–899 (cit. on p. 31).

[Yan20] YANG, Dongseok; LEE, Kanghee and CHOI, Younggeun: "TapSix: A Palm-Worn Glove with a Low-Cost Camera Sensor that Turns a Tactile Surface into a Six-Key Chorded Keyboard by Detection Finger Taps". In: *International Journal of Human–Computer Interaction* 36.1 (2020), pp. 1–14. DOI: 10.1080/10447318.2019.1597573. URL: https://doi.org/10.1080/10447318.2019.1597573 (cit. on p. 19).

[Yee07a] YEE, Nick and BAILENSON, Jeremy: "The Proteus Effect: The Effect of Transformed Self-Representation on Behavior". In: *Human Communication Research* 33.3 (2007), pp. 271–290. DOI: 10.1111/j.1468-2958.2007.00299.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-2958.2007.00299.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-2958.2007.00299.x (cit. on p. 13).

[Yee07b] YEE, Nick; BAILENSON, Jeremy N.; URBANEK, Mark; CHANG, Francis and MERGET, Dan: "The Unbearable Likeness of Being Digital: The Persistence of Nonverbal Social Norms in Online Virtual Environments". In: *CyberPsychology & Behavior* 10.1 (2007). PMID: 17305457, pp. 115–121. DOI: 10.1089/cpb.2006.9984 (cit. on p. 30).

[Yu17] YU, Chun; GU, Yizheng; YANG, Zhican; YI, Xin; LUO, Hengliang and SHI, Yuanchun: "Tap, Dwell or Gesture?: Exploring Head-Based Text Entry Techniques for HMDs". In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems.* ACM. 2017, pp. 4479–4488 (cit. on pp. 14, 39, 40).

[Yua10] Yuan, Y. and Steed, A.: “Is the rubber hand illusion induced by immersive virtual reality?” In: *2010 IEEE Virtual Reality Conference (VR).* 2010, pp. 95–102. doi: 10.1109/VR.2010.5444807 (cit. on p. 13).

[Yue74] Yuen, Karen K.: “The two-sample trimmed t for unequal population variances”. In: *Biometrika* 61.1 (Apr. 1974), pp. 165–170. doi: 10.1093/biomet/61.1.165 (cit. on pp. 55, 298).

[Zha16a] Zhang, Lixing and Matsumaru, Takafumi: “Near-Field Touch Interface Using Time-of-Flight Camera”. In: *Journal of Robotics and Mechatronics* 28.5 (2016), pp. 759–775. doi: 10.20965/jrm.2016.p0759 (cit. on p. 18).

[Zha16b] Zhang, Yang; Zhou, Junhan; Laput, Gierad and Harrison, Chris: “SkinTrack: Using the Body as an Electrical Waveguide for Continuous Finger Tracking on the Skin”. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.* CHI ’16. San Jose, California, USA: Association for Computing Machinery, 2016, pp. 1491–1503. doi: 10.1145/2858036.2858082. url: https://doi.org/10.1145/2858036.2858082 (cit. on p. 19).

[Zha18] Zhao, Yiwei and Follmer, Sean: “A Functional Optimization Based Approach for Continuous 3D Retargeted Touch of Arbitrary, Complex Boundaries in Haptic Virtual Reality”. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.* ACM. 2018, p. 544 (cit. on p. 31).

[Zha19] Zhang, Yang; Kienzle, Wolf; Ma, Yanjun; Ng, Shiu S.; Benko, Hrvoje and Harrison, Chris: “ActiTouch: Robust Touch Detection for On-Skin AR/VR Interfaces”. In: *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology.* UIST ’19. New Orleans, LA, USA: Association for Computing Machinery, 2019, pp. 1151–1159. doi: 10.1145/3332165.3347869. url: https://doi.org/10.1145/3332165.3347869 (cit. on pp. 16, 19).

[Zha20] Zhao, L.; Liu, Y.; Ye, D.; Ma, Z. and Song, W.: “Implementation and Evaluation of Touch-based Interaction Using Electrovibration Haptic Feedback in Virtual Environments”. In: *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR).* 2020, pp. 239–247. doi: 10.1109/VR46266.2020.00043 (cit. on pp. 18, 19).

[Zha96] Zhai, Shumin; Milgram, Paul and Buxton, William: “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.* ACM. 1996, pp. 308–315 (cit. on p. 14).

[Zhe17] Zheng, Zhongyang; Wang, Bo; Wang, Yakun; Yang, Shuang; Dong, Zhongqian; Yi, Tianyang; Choi, Cyrus; Chang, Emily J. and Chang, Edward Y.: “Aristo: An Augmented Reality Platform for Immersion and Interactivity”. In: *Proceedings of the 2017 ACM on Multimedia Conference - MM ’17.* The 2017 ACM. Mountain View, California, USA: ACM Press, 2017, pp. 690–698. doi: 10.1145/3123266.3123308 (cit. on p. 96).

[Zho20] Zhou, Qian; Sykes, Sarah; Fels, Sidney and Kin, Kenrick: “Gripmarks: Using Hand Grips to Transform In-Hand Objects into Mixed Reality Input”. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* CHI ’20. Honolulu, HI, USA: Association for Computing Machinery, 2020, pp. 1–11. doi: 10.1145/3313831.3376313. url: https://doi.org/10.1145/3313831.3376313 (cit. on p. 19).

[Zib18] Zibrek, Katja; Kokkinara, Elena and McDonnell, Rachel: “The Effect of Realistic Appearance of Virtual Characters in Immersive Environments - Does the Character’s Personality Play a Role?” In: *IEEE Transactions on Visualization and Computer Graphics* 24.4 (2018), pp. 1681–1690 (cit. on pp. 32, 135).

[Zie19] Zielasko, D.; Krüger, M.; Weyers, B. and Kuhlen, T. W.: “Passive Haptic Menus for Desk-Based and HMD-Projected Virtual Reality”. In: *2019 IEEE 5th Workshop on Everyday Virtual Reality*

(WEVR). Mar. 2019, pp. 1–6. DOI: 10.1109/WEVR.2019.8809589 (cit. on p. 18).

[Zim93] ZIMMERMAN, Donald W. and ZUMBO, Bruno D.: “Relative Power of the Wilcoxon Test, the Friedman Test, and Repeated-Measures ANOVA on Ranks”. In: *The Journal of Experimental Education* 62.1 (1993), pp. 75–86. DOI: 10.1080/00220973.1993.9943832. URL: https://doi.org/10.1080/00220973.1993.9943832 (cit. on p. 296).

[Živ03] ŽIVIČNJAK, Miroslav; SMOLEJ NARANČIĆ, Nina; SZIROVICZA, Lajos; FRANKE, Doris; HRENOVIĆ, Jasna and BIŠOF, Vesna: “Gender-Specific Growth Patterns for Stature, Sitting Height and Limbs Length in Croatian Children and Youth (3 to 18 Years of Age)”. In: *Collegium Antropologicum* 27.1 (2003), pp. 321–334 (cit. on p. 135).

List of Figures

List of Tables

Acronyms

2D	two-dimensional
3D	three-dimensional
AR	augmented reality
CAVE	cave automatic virtual environment
CVE	collaborative virtual environment
DOF	degrees of freedom
FOV	field of view
FPS	frames per second
GDI	Microsoft Windows graphics device interface
GPU	graphics processing unit
GUI	graphical user interface

HHI	human-to-human interaction
HIS	hybrid interaction system
HMD	head-mounted display
IMU	inertial measurement unit
PC	personal computer
PP	personal perspective
SfM	structure from motion
SLAM	simultaneous localization and mapping
UI	user interface
VE	virtual environment
VR	virtual reality
WIMP	windows, icons, menus and, pointer

A Questionnaires and Statistics for the User Studies

The ideas and implementations of this thesis are evaluated in several user studies as this allows to quantify user's subjective experience. To capture the workload, usability, or social interaction of a task, the following questionnaires are used:

Part one of the NASA-TLX or raw TLX (RTLX) [Har06] evaluates the perceived workload of the task and consists of the six scales mental demand, physical demand, temporal demand, performance, effort, and frustration level. These six scales are rated on a 21-point Likert scale from 0 (best value, i.e., very low/perfect) to 100 (worst value, i.e., very high/failure). The unweighted average of the ratings yields an overall score. The RTLX overall score is assessed on a scale from 0 (low workload) to 100 (high workload).

The User Experience Questionnaire (UEQ) [Sch14] assesses the usability and experience of a system. Users rate their experience on 26 scales of opposing pairs on a seven-point Likert scale. The results are combined to derive the six scores of attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty that are rated on a scale from -3 (bad) to 3 (excellent). To gain quick insights, a short version of the UEQ, the UEQ-S [Sch17], was developed. It measures the pragmatic, hedonic, and overall quality of a system using eight opposing pairs on a seven-point Likert scale. Again, the resulting scores range from -3 (bad) to 3 (excellent).

The Networked Minds Social Presence Measure (NMSPM) [Har04] quantifies the social presence of the interaction with 36 questions on six sub-scales. For the user studies, only the sub-scales co-presence, attentional allocation, perceived message understanding, and perceived behavioral interdependence were used and evaluated on a seven-point Likert scale. The scales perceived emotional understanding and perceived emotional interdependence were omitted because emotions and attitude can only be transferred to a limited extent with avatars that lack facial expressions. The six questions of each sub-scale of the NMSPM were aggregated to a total score for each of them. To make sure that the scales of the NMSPM are consistently rated as larger values are better, the ratings on the following scales of the questionnaire were inverted (as in [Piu18, Piu19]): Attentional allocation questions 1, 2, 5, and 6. Perceived message understanding questions 5 and 6. The resulting scales range from 1 (bad) to 7 (good).

At last, users were presented with custom questionnaires and asked to provide written feedback depending on the system which is explained in the respective sections.

The results of the evaluations are reported as average value (M) with standard deviation (SD), 95% confidence interval (CI), and root-mean-square error (RMSE). Most results are depicted as figures showing bar plots. However, some results show Box-Whisker-Plots [Bec16] and probability density functions using a kernel density estimation (KDE) [Lan04]. The bandwidth of the KDE is calculated using the Scott-rule [Sco12].

This thesis uses several significance tests to show that a difference among the mean values of the sampled data in the user studies also exists in the population of the data. The book *Modern Statistical Methods for HCI* [Rob16] provides a good overview of the commonly used satistical tests and gives practical examples on how to calculate the tests using the programming language *R*[1]. However, for most of the calculations in this thesis, the software *SPSS* was used[2].

The result of the calculation of a significance test is a test value, e.g., F-value for ANOVA. Based on this test value, a probability value (p) can be derived. To evaluate whether an effect is present, e.g., the means of two data sets are different, a null hypothesis is formulated that claims the effect is not present. If the p-value is smaller than a significance level of α, then the null hypothesis can be rejected and the original hypothesis is accepted, e.g., the difference

---

[1] https://www.r-project.org/

[2] https://www.ibm.com/analytics/spss-statistics-software/

between the means is not a random effect of the evaluation but exists in the population. If the p-value is larger than α, the null hypothesis cannot be rejected, e.g., there is no significant difference within the whole population. The probability value of $1 - \alpha$ describes the correct rejection of a null hypothesis. However, with a probability of α, a true null hypothesis is falsely rejected (type I error). Also, with a probability of β, a null hypothesis is accepted even though it should have been rejected (type II error). The value of $1 - \beta$ is equal to the power of a test. In this thesis, the null hypothesis is evaluated under the assumption of a significance level of α=.05. For all tests, the test value, the probability value (p), and also the size of the test set (N) are reported. Plots indicate significant differences with * for p≤.05, ** for p≤.01, and *** for p≤.001.

While performing an evaluation, different variables are captured. These can be grouped into independent and dependent variables. *Independent variables* describe *factors* that are changed during an experiment, e.g., algorithms that are compared. For each factor, there exist two or more *levels*, e.g., the factor of *algorithm* may have three levels *A*, *B*, and *C*. *Dependent variables* are influenced by these independent variables, e.g., users' task performance may be increased/reduced depending on the currently used algorithm. Each factor can affect the dependent variables (main effect). However, when two or more factors are present, the factors can also interact with each other (interaction effect), e.g., the interaction of the factors algorithm*device could show a certain device complements a specific algorithm.

Furthermore, there exist two test designs to execute an evaluation: A *between-subject* design uses independent samples of data recordings. This means that the measurements of a dependent variable within one level cannot be matched to another level, e.g., the users of algorithm *A* and the users of algorithm *B* are not the same. In contrast, a *within-subject* design uses dependent samples so that measurements can be matched between the levels, e.g., all users performed algorithms *A* and *B*.

Also, the distribution and level of measurement, i.e., nominal, ordinal, interval or ratio scale, of the collected data plays a role. *Parametric* tests often assume that the data fits a normal distribution and is interval or ratio scaled whereas *nonparametric* tests are distribution-free and can handle data that is at least of ordinal scale. Normal distribution can be tested using the Shapiro-Wilk test [Sha65]. For results of p>.05, the data samples are normally distributed. However, one can assume normality for larger sample sizes (N>20) and t-test and ANOVA are fairly robust against moderate deviations from normality.

Significance tests demand different requirements for the data, e.g., evaluation design, the number of factors and levels, distribution of samples, or level of measurement. Therefore, depending on the underlying data, different significance tests have to be used. Fig. A.1 gives an overview of commonly employed tests sorted by main requirements. In this thesis, the following tests are used: Paired t-test [Hsu05], Wilcoxon signed-rank (WSR) test [Wil70], Mann-Whitney-U-Test [Man47], one-way and two-way repeated-measures analysis of variance (ANOVA) [Fis92], and Friedman test [Zim93].

The ANOVA test assumes sphericity of the data. This means that the relationship between levels is similar and their variances are approximately equal. For two levels, sphericity is always met. To measure if sphericity is met for more than two levels, Mauchly's test can be used. If the p-value of Mauchly's test is not significant, the data can be assumed to not violate the assumption of sphericity. However, if Mauchly's p-value is significant, then the null hypothesis has to be rejected which indicates a violation of sphericity. In that case, the degrees of freedom of the ANOVA have to be adjusted by a correction term ϵ. The ϵ of Greenhouse-Geisser is more restrictive and is used if Greenhouse-Geisser $\epsilon < .75$. If Greenhouse-Geisser $\epsilon > .75$, then the Huynh-Feldt ϵ is used for correction.

When there are more than two levels, different post hoc tests are applied to check for significant differences between conditions. This is necessary because performing pairwise comparisons accumulates probability of α and leads to an inflation of a type I error. In this thesis, pairwise comparisons are calculated using Bonferroni-corrected [Bon35] p-values and Tukey HSD post hoc tests [Tuk53]. When there are few pairwise comparisons, Bonferroni correction provides higher statistical power, and for many pairwise comparisons, a Tukey test has more power. Bonferroni correction is calculated using $p^* = p/N$. The Tukey tests are adjusted with a family-wise error rate (FWER) of 1% to avoid the inflation of a type-1 error.

Besides checking data for statistical differences, also equivalence tests are

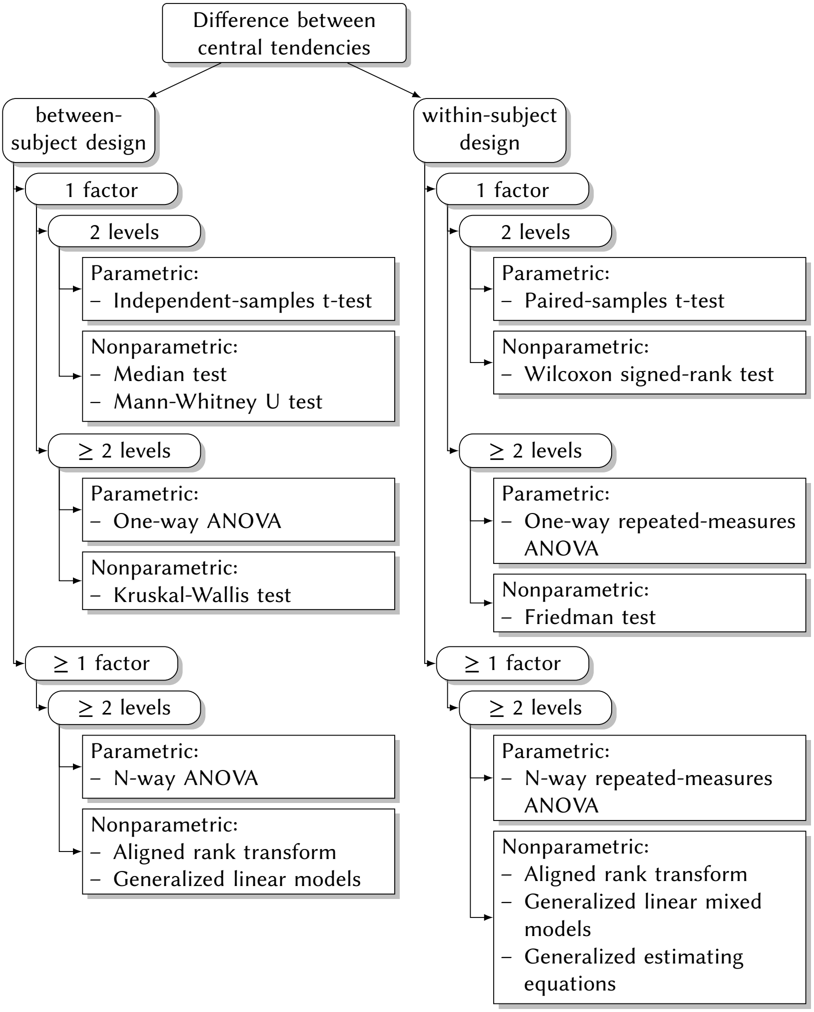


Figure A.1: Overview of significance tests to analyze the differences among means grouped by study design, factors, and levels based on [Rob16] (p. 138).

used to show the equality of levels. This thesis only uses a robust two one-sided test (rTOST) [Yue74] which applies to within-subject designs with one factor and two levels.

As described, the p-value indicates whether an effect within the results of the evaluation is also applicable to the population or not. If an effect is significant, its strength can be measured using effect size. In this thesis, effect sizes are reported with Cohen's d [Coh88], Pearson's correlation coefficient r [Pea31], partial eta squared (η^2) [Coh73], or Kendall's coefficient of concordance W [Win40]. The effect sizes are chosen based on the resulting values of the significance or equivalence test. However, many effect sizes can be converted to other types. Cohen's d is reported for rTOST tests, Pearson's r for WSR and t-tests, partial η^2 for ANOVA, and Kendall's W for Friedman tests. Table A.1 shows how the different values of the effect sizes can be interpreted.

Table A.1: Interpretation of effect sizes

	Cohen's d	**Pearson's r**	**partial η^2**	**Kendall's W**
small	≥.2	≥.10	≥.01	≥.1
medium	≥.5	≥.25	≥.06	≥.3
large	≥.8	≥.40	≥.14	≥.5