Enable Spatial Interaction for Distant Displays for Everyone

Doctoral thesis

for obtaining the academic degree

Doctor of Natural Sciences (Dr. rer. nat.)

submitted by

Teo Babic

at the

Universität
Konstanz

Faculty of Sciences

Department of Computer and Information Science

Konstanz, 2024

Teo Babic
Enable Spatial Interaction for Distant Displays for Everyone. Doctoral thesis for obtaining the academic degree Doctor of Natural Sciences (Dr. rer. nat.).

University of Konstanz
Human-Computer Interaction Group (HCI Group)
Faculty of Sciences
Department of Computer and Information Science
Universitätsstraße 10
78457 Konstanz

Date of the oral examination: May 06, 2024
1. Reviewer: Professor Dr. Harald Reiterer (University of Konstanz)
2. Reviewer: Professor Dr. Michael Haller (Free University of Bozen-Bolzano)

Abstract

In the evolving landscape of Human-Computer Interaction (HCI), this thesis confronts a pivotal question: Is it possible to design a spatial interaction system for distant displays that is accessible and usable by everyone, without the need for dedicated hardware? This inquiry propels the research into exploring the potential of smartphones, repurposing them as efficient controllers and trackers for spatial interaction. The thesis begins with a thorough analysis of existing spatial interaction systems, identifying design challenges and opportunities, and setting the stage for innovative solutions.

Embarking on a methodical journey, the thesis is structured into distinct objectives. The first objective analyzes current spatial interaction systems, uncovering their limitations and potentials. Building on these insights, the second objective is the development and implementation of a new smartphone-based tracking solution, tailored for enhancing spatial interactions with distant displays. This venture comprises designing multimodal interaction techniques, establishing a robust and comprehensive interaction system, and empirically evaluating the developed solutions. Multiple studies delve into how smartphones, equipped with advanced sensors and cameras, can facilitate multimodal interactions ranging from touch and hand gestures to head and eye movements. The third objective generalizes the findings, transforming them into practical guidelines and applications for designers and developers. This encompasses in the discussion and conclusion of the research, and offers a roadmap for future developments in spatial interaction using personal devices.

Conclusively, this thesis not only demonstrates the feasibility of leveraging everyday smartphones for sophisticated spatial interaction but also paves the way for more inclusive and universally accessible interaction systems. It signifies a paradigm shift in HCI, where the ubiquity and versatility of personal devices are harnessed to transcend the boundaries of traditional spatial interaction methods.

Kurzfassung

In der sich entwickelnden Landschaft der Mensch-Computer-Interaktion konfrontiert diese Dissertation eine zentrale Frage: Ist es möglich, ein räumliches Interaktionssystem für entfernte Displays zu entwickeln, das für jeden zugänglich und nutzbar ist, ohne spezielle zusatz Hardware zu benötigen? Diese Untersuchung treibt die Forschung voran, um das Potenzial von Smartphones zu erkunden, die als effiziente Controller und Tracker für räumliche Interaktionen neu eingesetzt werden. Die Arbeit beginnt mit einer gründlichen Analyse bestehender räumlicher Interaktionssysteme, identifiziert Herausforderungen und Chancen und bereitet den Weg für innovative Lösungen.

Auf einem methodischen Weg ist die Dissertation in spezifische Ziele gegliedert. Das erste Ziel analysiert aktuelle räumliche Interaktionssysteme und deckt ihre Grenzen und Potenziale auf. Aufbauend auf diesen Erkenntnissen ist das zweite Ziel die Entwicklung und Implementierung einer neuen smartphonebasierten Tracking-Lösung, die für die Verbesserung der räumlichen Interaktion mit entfernten Displays maßgeschneidert ist. Dieses Unterfangen umfasst das Design multimodaler Interaktionstechniken, die Etablierung eines robusten und umfassenden Interaktionssystems und die empirische Bewertung der entwickelten Lösungen. Mehrere Studien untersuchen, wie Smartphones mit fortschrittlichen Sensoren und Kameras multimodale Interaktionen ermöglichen können, von Touch- und Handgesten bis hin zu Kopf- und Augenbewegungen. Das dritte Ziel verallgemeinert die Ergebnisse und wandelt sie in praktische Richtlinien und Anwendungen für Designer und Entwickler um. Dies umfasst die Diskussion und Schlussfolgerungen der Forschung und bietet einen Fahrplan für zukünftige Entwicklungen in der räumlichen Interaktion mit persönlichen Geräten.

Zusammenfassend zeigt diese Dissertation nicht nur die Machbarkeit, alltägliche Smartphones für anspruchsvolle räumliche Interaktionen zu nutzen, sondern ebnet auch den Weg für inklusivere und universell zugängliche Interaktionssysteme. Sie kennzeichnet einen Paradigmenwechsel in der Mensch-Computer-Interaktion, bei dem die Allgegenwart und Vielseitigkeit persönlicher Geräte genutzt wird, um die Grenzen traditioneller räumlicher Interaktionsmethoden zu überschreiten.

Acknowledgements

This thesis would not have been possible without the help and support of many people. First, I would like to thank my secondary advisor, Prof. Dr. Michael Haller, who introduced me to HCI and academic pursuits prior to this thesis and continuously inspired me with his enthusiasm for a wide spectrum within computer science. Furthermore, I would like to thank my primary advisor, Prof. Dr. Harald Reiterer, for placing your trust in me and supporting my work throughout this thesis journey. You both always made it possible for me to work self-determined and to pursue my ideas and research interests. At the same time, you always had an open door, were pleased to give advice, and stood behind me if problems occurred. I am very gratified that I have been able to work with the HCI Group of the University of Konstanz as well as the Media Interaction Lab in Hagenberg for many years. Their environments have been immensely conducive to my growth and development. Additionally, I sincerely thank Dr. Felix Schwarz for his support and guidance in my research activities within the BMW Group industry Promotion programme.

Being surrounded by remarkable and inspiring individuals greatly contributed to this work, laying the groundwork for my growth and development. Special thanks to the Media Interaction Lab Team in Hagenberg - Christian, Florian, Joanne, Kathrin, and Patrick - for your initial support while I was embarking on this PhD journey. Your expertise and professionalism were key to my personal growth. Being part of this wonderful team has been an enriching experience, and I hope our paths intertwine again in the future. Additionally, my appreciation goes to the PhD HCI team at the BMW Group - Florian, Jeremy, Yannick, Stefan H., Stefan B. - for your valuable discussions, unwavering support, and motivation throughout our collective PhD endeavors. Your cheerful nature and support during challenging times have been a constant source of encouragement.

Above all, I am indebted to my lovely wife, Urška. I am deeply grateful for your endless love, your belief in me, your unfailing support, and patience, for all the late nights and weekends you accepted to spend without me, and for the cheerful and restful moments, you ensured even during the tough times of writing this thesis. Thank you for accompanying me to all this.

Related Publications

Parts of the publications listed below are presented in this thesis. Reused material is highlighted in the introduction and in the beginning of the corresponding chapters. For a better overview, publications that contributed to the core of this thesis are listed first. Other publications that are not directly part of this dissertation are listed second. They are presented chronologically, starting with the oldest publication:

Publication Included in this Thesis

Teo Babic, Harald Reiterer, and Michael Haller. *"Pocket6: A 6DoF Controller Based On A Simple Smartphone Application"* In: Proceedings of the Symposium on Spatial User Interaction - SUI '18. New York, New York, USA: ACM Press, 2018, pp. 2–10. https://doi.org/10.1145/3267782.3267785 *Chapter 3*

Received an Honorable Mention Award as the 2nd best paper on conference.

Teo Babic, Florian Perteneder, Harald Reiterer, and Michael Haller. *"Simo: Interactions with Distant Displays by Smartphones with Simultaneous Face and World Tracking"* In: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems - CHI EA '20. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3334480.3382962 *Chapter 4 and 5*

Teo Babic, Harald Reiterer, and Michael Haller. *"Understanding and Creating Spatial Interactions with Distant Displays Enabled by Unmodified Off-The-Shelf Smartphones"* In: Multimodal Technologies and Interaction. 2022; 6(10):94. https://doi.org/10.3390/mti6100094 *Chapter 1 and 5*

Additional Publications

Publication Excluded from this Thesis

Teo Babic, Harald Reiterer, and Michael Haller. *“GestureDrawer: One-handed Interaction Technique for Spatial User-defined Imaginary Interfaces”* In: Proceedings of the 5th Symposium on Spatial User Interaction - SUI ’17. New York, New York, USA: ACM Press, 2017, pp. 128–137. https://doi.org/10.1145/3131277.3132185

Teo Babic, Harald Reiterer, and Michael Haller. *“GestureDrawer Demo: One-handed Interaction Technique for Spatial User-defined Imaginary Interfaces”* In: Proceedings of the 5th Symposium on Spatial User Interaction - SUI ’17. New York, New York, USA: ACM Press, 2017, pp. 149–149. https://doi.org/10.1145/3131277.3134363

Michael Julian Kronester, **Teo Babic**, Andreas Riener and Simon Nestler. *“Conceptual Design and Evaluation of Vibrotactile Feedback on the Wrist to Enrich the Perception of Virtual Objects Whilst Performing Mid-Air Gestures”*. Technische Hochschule Ingolstadt, Ingolstadt, 2020. http://nbn-resolving.de/urn:nbn:de:bvb:573-8219

Michael Julian Kronester, Andreas Riener, and **Teo Babic**. *“Potential of Wrist-worn Vibrotactile Feedback to Enhance the Perception of Virtual Objects during Mid-air Gestures”* In: Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems - CHI EA ’21. Association for Computing Machinery, New York, NY, USA, Article 256, 1–7. https://doi.org/10.1145/3411763.3451655

Patents

Teo Babic. *“Car Part Selection and Manipulation Framework for Wearable and Handheld Devices”*. Patent No. 2017218780 PFF, Filed July 1st, Issued Aug. 9th., 2018. https://patents.google.com/patent/DE102017218780A1

Contents

> *“The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it.”*
>
> *Mark Weiser*

1

Introduction

This thesis delves into interaction techniques designed to augment a user's ability to engage with computers in innovative and efficient manners. The primary focus is on a smartphone-based multimodal interface, developed by integrating motion tracking (including hand gestures, head pointing, and eye gaze) with inputs like touch in diverse ways. This exploration aims to comprehend the interface's design and usability in common human-computer interaction contexts. We focus on spatial interaction techniques applicable to both 2D and 3D user interfaces and content. Specifically, we introduce a novel concept enabling ordinary smartphone users to expand their typically restricted interactive capabilities with content shown on a distant display. Imagine a user wanting to interact with a displayed 3D model. Traditional spatial interaction relies on specialized tracking devices, which are often expensive and not readily available. This presents a barrier for many users who might not have access to such tools. With the system proposed in this thesis, users can use their own smartphone to manipulate and delve into the 3D model, breaking down the barriers of specialized equipment and democratizing access to spatial interaction techniques. This thesis proposes solutions to overcome these limitations.

In the remainder of this chapter, we firstly introduce the research context and motivation of this thesis. Next, we establish the research scope and the research objectives that this thesis follows to tackle the main research question. Finally, we provide an overview of the methodical approach, point out the main contributions of this thesis, and give a preview of the upcoming chapters.

1

Introduction

This thesis delves into interaction techniques designed to augment a user's ability to engage with computers in innovative and efficient manners. The primary focus is on a smartphone-based multimodal interface, developed by integrating motion tracking (including hand gestures, head pointing, and eye gaze) with inputs like touch in diverse ways. This exploration aims to comprehend the interface's design and usability in common human-computer interaction contexts. We focus on spatial interaction techniques applicable to both 2D and 3D user interfaces and content. Specifically, we introduce a novel concept enabling ordinary smartphone users to expand their typically restricted interactive capabilities with content shown on a distant display. Imagine a user wanting to interact with a displayed 3D model. Traditional spatial interaction relies on specialized tracking devices, which are often expensive and not readily available. This presents a barrier for many users who might not have access to such tools. With the system proposed in this thesis, users can use their own smartphone to manipulate and delve into the 3D model, breaking down the barriers of specialized equipment and democratizing access to spatial interaction techniques. This thesis proposes solutions to overcome these limitations.

In the remainder of this chapter, we firstly introduce the research context and motivation of this thesis. Next, we establish the research scope and the research objectives that this thesis follows to tackle the main research question. Finally, we provide an overview of the methodical approach, point out the main contributions of this thesis, and give a preview of the upcoming chapters.

1.1 Research Context

The research presented in this thesis is situated within the broader context of *Human-Computer Interaction* (*HCI*), a multidisciplinary field that addresses the design and use of computer technology, with a particular emphasis on interfaces and interactions between humans and computers. HCI incorporates elements from computer science, behavioral sciences, design, media studies and several other academic disciplines. This thesis discusses topics within the broad domain of *Spatial Interaction*, *Interaction Techniques*, and *Interactive Devices*, all of which fall under the domain of HCI.

The focus of this thesis is further refined to the domain of *spatial interaction*, a specialized subfield within HCI. *Spatial interaction* can be understood as the way users engage with and navigate through virtual or augmented reality environments, 3D interfaces, and spatial computing applications. It concerns how input like body motions, gestures, voice, or traditional input devices can be translated into movement and control within a three-dimensional space, as well as how feedback is provided to the user. The precise definition of *spatial interaction* may vary depending on the specific area within HCI, for example, between areas such as virtual reality, augmented reality, or 3D user interfaces. Additionally, *spatial interaction* shares numerous similarities and overlaps with other related disciplines. One notable field with significant overlap is *3D user interfaces* (*3DUI*). For instance, the book *"3D User Interfaces: Theory and Practice"* by LaViola et al. provides a comprehensive overview of 3D user interfaces, covering both theoretical and practical aspects of spatial interaction within virtual environments. The book provides insights into design principles, interaction techniques, and various applications of 3D interfaces [144]. Another field that intersects with *spatial interaction* is the field of *natural interfaces*. In the book *"Brave NUI"* Wigdor et al. provide insights into designing *natural user interfaces* (*NUI*), which often incorporate spatial elements [284]. It offers a thorough examination of touch and gesture-based interaction, discussing the challenges and considerations in designing interfaces that respond to users' movements in 2D and 3D space.

In particular, our research explores spatial interaction in the context of distant displays. *Distant displays* are screens or projections that are not within arm's reach of the user, making traditional touch-based interaction not possible. Examples include large public displays, home television screens positioned at a distance from the user, or even virtual displays in augmented or virtual reality environments. *Spatial interaction with distant displays* involves techniques and strategies allowing users to engage with out-of-reach displays, ensuring effective input and feedback despite physical distances. These interactions typically require a combination of input devices, gestures, and gaze tracking, enabling users to navigate, select, and manipulate content on distant displays. Considering the diverse environments where distant displays are utilized — such as homes, public areas, and workplaces — the scope of our investigation is intentionally narrowed to concentrate on spatial interactions with distant displays. In the paper "Distant freehand pointing and clicking on very large, high-resolution displays" by Vogel et al., the authors explore the challenges and solutions associated with the interaction, more specifically freehand pointing and clicking, on distant displays [268]. They examine the feasibility and accuracy of freehand interaction techniques, where users employ hand gestures in the air (without any physical device or touch interaction) to control and manipulate content on distant displays. Various interaction techniques are proposed and assessed by the authors, incorporating visual feedback and control-display gain adaptation to improve the accuracy and user experience associated with distant freehand pointing and clicking. This paper is an exemplary study in the field, illustrating the complexity and considerations involved in facilitating user interaction with distant displays through spatial input techniques.

1.2 Motivation

This chapter is structured into two subsections to discuss the thesis's motivation. Subsection 1.2.1 starts with a historical perspective and later introduces the imperative concept of universal design in computing interfaces, emphasizing the transition from exclusive to inclusive access to computer technologies. It lays the groundwork by exploring the term *universal design*, a principle vital for creating products accessible to every individual, regardless of age, ability, or culture. Subsection 1.2.2 then narrows the focus to spatial interaction with distant displays, providing insights into the specific challenges and opportunities in this domain. It aims to deepen the understanding of making spatial interaction technologies universally accessible and usable, setting the stage for subsequent discussions in the thesis.

1.2.1 Designing for Everyone

Over the past fifty years, the innovation of computing devices has significantly influenced the progress of humanity. Collectively, connected computers of various forms in the hands of so many people changed the way we live, work, socialize, and entertain ourselves. There are hardly any tools left that would not rely on some sort of computational device with a user-facing interface used for its operation. Even if a tool does not directly incorporate a computing device, it is highly likely that computers played a role in its development process, from ideation and design to construction and production. A popular observation describing these advancements in digital electronics is Moore's Law, which states that already since the 1960s, we can expect the speed and capability of our computers to increase every couple of years, and that we will pay less for them. These rapid advancements in digital electronics have propelled technological and social change, boosting productivity and economic growth. However, while Moore's Law outlines the technical advancements in computational hardware, it does not address the continuous user demand for these devices, which in turn fuels the cycle described by Moore's Law. Ben Shneiderman, one of the pioneers in the field of human-computer interaction, emphasized in his CHI 2017 opening keynote that the relentless pursuit of inclusivity and diversity in user experience has been a driving force behind the market demand that sustains Moore's Law [235]:

> “ *We created the market pull for Moore's Law, since we believe in the diversity of users, and we wanted to include everyone.* ”
>
> *Ben Shneiderman*

The quote refers to the early days of computers when only a few highly educated individuals had access to computers and had the know-how to operate them. This exposes that there are some key aspects on which we need to work as human-computer interaction scientists so that new computer technology can be adapted by the masses and that its potential can be used on a large scale.

To describe the concept of designing products to be used to the greatest extent possible by everyone, regardless of their age, ability, or culture, we can refer to the term and idea of *universal design*. Initially, this term comes from architecture [253], where it presented the motivation behind solutions as barrier-free design. Subsequently, the concept was embraced in the field of human-computer interaction, supporting early efforts and solutions designed to make computer systems accessible also for users with physical disabilities [251, 252].

In recent years, we can note that the field of human-computer interaction focuses more than ever on universal design or one of its subfields as, for example, *inclusive design*. The expansion of these subfields has led to the identification of various specific user groups for whom designers can tailor their systems. Systems are now being developed with consideration for users with disabilities (encompassing auditory, cognitive, physical, and visual impairments), different language groups, various age demographics (from the elderly to children), individuals with cultural or ethnological differences, users of differing physical sizes, individuals who wear glasses or use assistive devices, and those who differ in *digital divide* — a term used to explain the gap between demographics and regions that have access to modern information and communications technology, and those that do not or have only restricted access. However, the proliferation of these subfields has led to the emergence and varied interpretation of terms like *universal design*, *inclusive design*, *accessibility*, and *user diversity*. There are, however, recent approaches that attempt to structure this field, with one notable example being the definition of *universal design* provided by Bruno Perez, grounded in the Fuji Xerox Guideline for Universal Design [195]. There, the authors define *universal design* as "The way we design products or environments to make them accessible and usable in the widest possible range of situations without the need for specific adaptation.". Furthermore, an even more interesting aspect of this guideline is the outlined three-step process to realize *universal design*, as depicted in Figure 1.1.

Another recent definition comes from the Nielsen Norman Group, which separated the terms in the following manner. *Accessibility* is focused on ensuring that interfaces and technology can be used by people with disabilities [81]. They point out that their definition of *accessibility* has a narrower scope than *inclusive design* in that it is focused on specific accommodations. They further note that this definition allows for a somewhat straightforward assessment of *accessibility* compared to *inclusive* and *universal design*. On the other hand, according to their viewpoint, the term *inclusive design* describes methodologies to create products that understand and enable users of all backgrounds and abilities. Various subfields within *inclusive design* address issues related to accessibility, age, culture, economic status, education, gender, geographic location, language, and race, with the objective

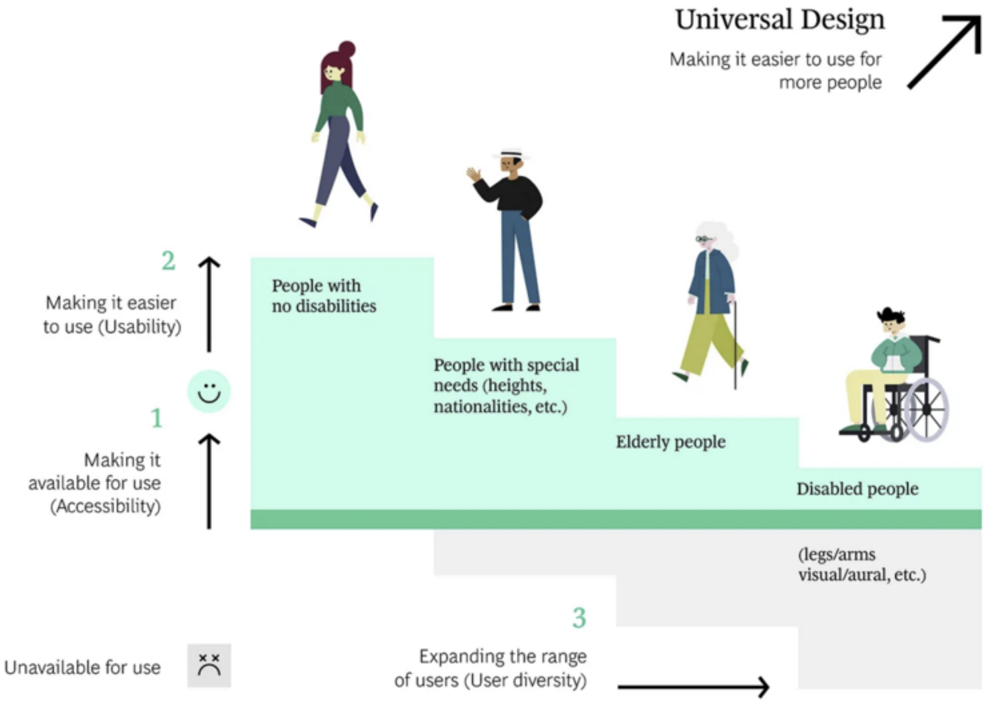


Figure 1.1: A definition of *universal design* including the sequential steps required to attain it. The progression towards *universal design* begins with achieving *accessibility*, then advancing to *usability*, and finally, embracing *user diversity*, based on [195].

.

being to meet as many user needs as possible, rather than merely accommodating as many users as possible. To zoom out even further, they define *universal design* as an aim to create one experience that can be accessed and used by the widest possible audience.

In the context of this thesis, it is important to define certain terms that will aid in establishing the objectives outlined in subsequent sections while also creating a foundation of shared terminology. Below are definitions of crucial aims to consider in pursuit of "design for everyone" when developing new interaction systems:

- **Universal Design** aims to make new technology to be used by as many users as possible. The general approach of making products used by as many users as possible, without designing for a specific group in mind. The objective is to tackle and overcome limitations that create limited access and make new technology unavailable for many users. To pursue *universal* *design*, we need to provide minimum ways to allow users to use a product, thereby making an experience open to everyone.

- **Usability** aims to make new technology systems easier to use for all its users by focusing on creating interaction principles that are more effective (improving the accuracy and completeness with which users achieve goals), efficient (minimizing the resources, like time, required to achieve goals) and increased user satisfaction (improving the comfort and acceptability of the system during use).
- **Application Flexibility** aims to integrate new technological solutions into a broad array of usage contexts or application scenarios. To achieve this, it is important that the interaction techniques incorporated within the new system, along with any potential usage limitations, do not hinder the technology's functionality in diverse scenarios. For example, usage while sitting or standing, under various lighting conditions or in applications that are either 2D or 3D.

When reflecting on the earlier examples regarding personal computers and the relation to Moore's Law, it is evident that the development of personal computers closely aligned with the three outlined aims above. Initially, computing devices were exclusive, accessible only within certain corporate or research environments due to their prohibitive costs, size, power consumption, and other means. With the adoption of principles today described with *universal design*, significant changes occurred in computers' performance, cost, and form factor, slowly democratizing access to these devices. This democratization allowed for widespread home ownership of computers, and today, similar computational devices such as laptops and smartphones accompany us everywhere. In parallel, alongside these changes, there was also a remarkable improvement in *usability*; while early computers demanded expertise with text-based commands, modern devices offer simple graphical interfaces navigable by users of all levels of expertise, even without specialized training. Finally, in terms of *application flexibility*, the role of computers has expanded from specialized task completion to serving a wide array of functions, supporting diverse applications in areas like productivity, communication, entertainment, gaming, and social networking, accompanied by ever-more standardized interaction techniques.

1.2.2 Spatial Interaction for Everyone

While personal computers such as laptops or smartphones are already heavily intertwined in the daily activity of billions of users, there are still newer technologies with new beneficial capabilities that are still inaccessible and unknown to many users. Here, the domain of human-computer interaction still needs to do its work on making such new devices to be used by as many users as possible. The area of *spatial interaction with distant displays*, the central topic of this thesis, is an area of such immaturity. In contrast to devices such as smartphones or laptops, only more technology affine users have access to devices allowing spatial human-computer interaction, while their applications are often only limited to specific domains, presenting interfaces and interaction methods that many find challenging.

To leverage the capabilities and benefits of spatial interaction today, users first need a device equipped with spatial user-motion tracking features. This requires additional dedicated hardware like computer vision sensors or cameras capable of capturing and tracking movements of different user's body parts. These user motions can then be translated into meaningful spatial interactions with the given application. However, the essential hardware for enabling these interactions is not available at no cost; users must purchase, maintain, and often install these devices in their space or even wear the device on their body.

Consider a scenario where a car exterior designer wishes to showcase a 3D rendering of a new car design to a team of ten colleagues. What immediate and cost-free alternatives are available for this presentation? Today still, the most common method is that the designer would employ a large 2D screen to display the 3D rendering application, manipulating the 3D model's viewpoint during the presentation using conventional input devices like a mouse and keyboard. Any alternative would likely necessitate financial investment, such as acquiring a motion-tracking system like the Microsoft Kinect sensor. This would allow the designer to manipulate the 3D model through hand-free gestures, tracked by the sensor while standing in front of the screen. More costly solutions might involve equipping the designer and perhaps all attendees with devices like Apple Vision Pro glasses, enabling shared viewing and discussion of the 3D model. From this example, we can see that there are not that many options that could be easily enabled and accessed, ideally cost-free and without specific technical setup or knowledge.

The overview provided in Figure 1.2 showcases a series of pivotal projects over a 50-year timeline. These projects, initially significant within the research community, later became examples of end-user devices for spatial interaction adopted by the consumer electronics industry. It is noteworthy that, despite the passage of decades, spatial interaction, whether with distant displays or within virtual/augmented reality environments, still fundamentally depends on devices specifically designed to enable spatial interactions. Engaging in spatial interactions appears neither feasible nor practical without these additional, purpose-built devices.

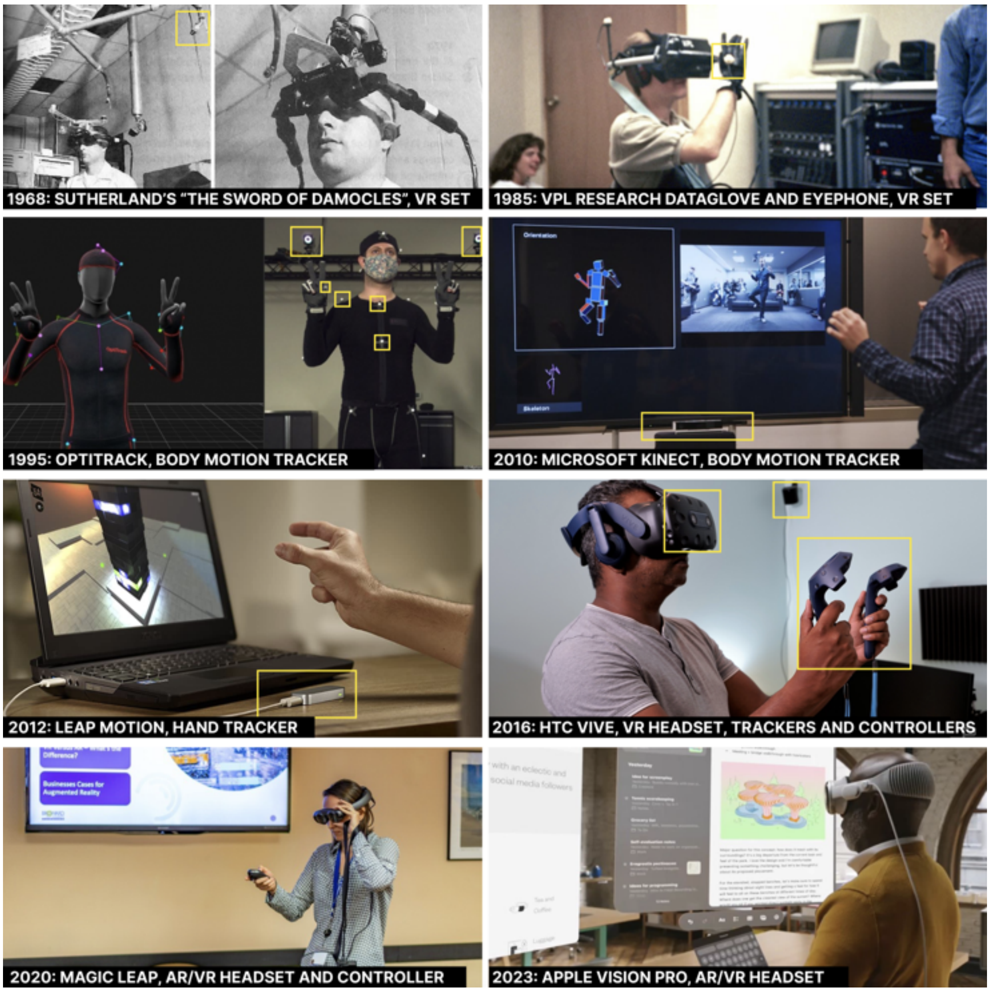


Figure 1.2: A timeline of the past 50 years, showing various devices to facilitate spatial computing, encompassing desktops, distant displays, and augmented and virtual reality. A commonality among these devices is the necessity to purchase and set up dedicated tracking devices before initiating any spatial interaction experience.

If we delve a bit deeper to understand why we have this case, we see there are multiple limitations that are hard to overcome in spatial interactive systems. Spatial interactive systems rely on some sort of spatial user motion tracking, by which the user can create the desired inputs or interactions. Present-day solutions, in terms of *usability*, suffer from a multitude of limitations that users must be aware of while using these spatial input devices. For instance, users are required to remain within the field of view of tracking cameras, thereby restricting their freedom of movement. Additionally, users are often compelled to execute hand or body gestures of a defined scale that is detectable by the tracking system, potentially leading to discomfort or physical muscle fatigue. There are also restrictions on users' mobility, as they cannot seamlessly transition between standing and sitting positions; often, they must adhere to and maintain a predetermined pose. Finally, when considering *application flexibility*, it is clear that the field of spatial interaction is fragmented by the types of hardware devices employed; the devices facilitating 2D desktop spatial applications differ from those used for distant display interactions, which in turn are again distinct from devices designed for virtual/augmented reality. Consequently, there is not a single device available that can be repurposed across various application areas, whether they involve 2D or 3D, distant display, or real or augmented/virtual reality environments. Furthermore, there is a conspicuous lack of significant initiatives aimed at developing a spatial interactive system that is more universal and standardized.

Design Challenge 1: Tracking Device Options

The challenge is determining whether the acquisition of dedicated devices is the sole viable option for users to effectively engage in and utilize spatial interactions with distant displays.

Despite their evident limitations, we believe that the devices currently available for enabling spatial interaction each individually demonstrate promising and beneficial approaches. However, what these approaches and devices critically lack are innovative strategies to significantly enhance their *universal design*, *usability*, and *application flexibility*, simultaneously, all at once. Together, these three aims could enable a broader user base to explore the potential of spatial interaction. It would be inaccurate to say that existing devices are unused, commercially unsuccessful, or that their applications are not beneficial. Nevertheless, we believe that a serious push still needs to be made to find ways that users could instantly gain the possibility to interact spatially in their daily lives without the need to buy additional hardware devices. Ideally, this could be achieved through a software-only approach, utilizing personal devices that users already possess, like smartphones. The envisioned scenario is one where users can access rich spatial interactions simply by downloading an app on their smartphones, eliminating the need to purchase, set up, and wear new devices. This approach would entirely eliminate the necessity for dedicated devices, potentially providing access to spatial interaction to millions of users. It is also worth noting that researchers and developers have made significant advancements in optimizing tracking capabilities, enhancing application usability, and beginning to establish interaction technique standards. However, we see a high risk that by eliminating the currently established technological foundation, by avoiding the use of dedicated hardware, we could compromise heavily on the previously mentioned achievements, optimizations, and established standards that we already made so far. For instance, interaction techniques that have recently begun to establish as standards for spatial interaction, such as gaze-pointing or hand-based ray-casting originating from virtual/augmented reality scenarios, might not be feasible in simpler systems without dedicated hardware. While such an approach would eliminate the need for dedicated hardware, it would represent a step back in terms of interaction techniques.

Design Challenge 2: Impact of Hardware Removal

The challenge is determining whether the elimination of dedicated hardware devices from spatial interaction systems unintentionally leads to a reduction in the effectiveness of established interaction techniques or impairs the overall user interaction experience.

We notice this in multiple attempts where researchers have designed systems with simplified hardware setups, often utilizing only smartphones. While they introduced valuable applications and effective solutions, they still needed to make significant compromises. Opting for simplified hardware approaches resulted in compromises regarding interaction techniques or the user interface. This means, the reduction in the capability or the plain number of motion sensors limits the range of feasible interaction techniques. In practice, many designers and researchers encounter this paradigm, for example, if they use a simple smartphone instead of an 8-camera OptiTrack lighthouse tracker, this directly limited the systems interaction possibilities since not that many body parts could be tracked. The common approach in the research community is to accept these limitations and design around the technical limitations of new interaction techniques based on the compromised tracking capabilities. In such situations, designers and researchers frequently find themselves unable to employ the most efficient or standardized interaction techniques. For example, consider eye-gaze tracking, a simple pointing method; if technical constraints prevent developers from implementing this feature, they are compelled to explore alternative approaches like hand-motion or touch-based pointing methods. Such decisions are driven by technical constraints rather than considerations of interaction design or usability. This often resulted in interactive systems that come with simple hardware but require a unique user interface or interaction technique. For instance, a device might incorporate a unique set of touch or hand-motion gestures, necessitating users to learn this specific language solely for operating that particular device. This illustrates how the complexity of spatial interactive systems can easily get transferred from one aspect to another. This situation prompts an exploration into the possibility of proving the contrary. Is it indeed accurate to say that more sophisticated hardware enables a broader range of interaction techniques for designers to use, while simpler hardware limits these interactions? Following this, a more compelling question emerges concerning the precise limitations: To what degree can one reduce dedicated hardware without substantially compromising interaction capabilities? This question projected on the context of spatial interaction with distant displays still remains unanswered.

Design Challenge 3: Personal Device Feasibility

The challenge is assessing the feasibility of using a personal device to enable spatial interactions and determining the extent of its spatial interactive capabilities.

In conclusion, the guiding motivational thought behind this thesis is to effortlessly enable as many users as possible for spatial interaction with distant displays via an already-owned personal smartphone while keeping the tracking capabilities already known from comparable solutions using dedicated hardware devices and while guaranteeing key usability characteristics from performance to user's motion freedom.

1.3 Research Objectives

Based on the correlating design challenges mentioned in the previous section, we can raise the question of whether an interaction system can be designed with a much better balance between device ease of accessibility, ease of use, and spatial interaction capabilities. Based on that, we set the main research question of this thesis:

Main Research Question of this Thesis

Can we build a spatial interaction system for distant displays that everyone can access and use without requiring a dedicated hardware device?

Answering this question is the main objective of this thesis. To do so, we present three research objectives. In the first, we investigate in detail the current solutions. In the second, including its corresponding sub-objectives, we create, implement, and test our newly proposed interactive system. Finally, in the third objective, we consolidate our findings for designers and developers. In the following, we formulate the *Research Objectives* of this thesis:

Objective 1: **Identifying drawbacks and challenges as well as opportunities of existing spatial interaction systems.** Interactive systems enabling spatial interactions are not novel device categories since they have been around for almost 30 years. Hence, in order to improve the usability and usefulness of this device category, the first objective is to analyze the status quo. This analysis will inform the next steps of this work.

Objective 2: **Investigating strategies to overcome the identified issues and further improve existing advantages by developing a new smartphone-based software solution for spatial interactions with distant displays.** Based on the identified issues and the opportunities identified in Objective 1, we plan to investigate possible solutions. The investigation includes (1) the interaction design of multimodal interaction techniques, (2) an implementation that comprehensively addresses the problem, (3) a thorough evaluation of the delivered solution, and (4) example applications that show the system's usage in various application scenarios. This objective aims to find new ways how to enable handheld-based spatial interaction without compromising tracking capability or application usability.

Objective 3: **Generalizing the collected findings in ways so that they can be helpful for designers and developers, who are creating applications and systems for spatial interaction.** The main goal of this objective is to find a way to present and communicate the findings from the previous objectives, which are not tightly connected to a specific use case or scenario. The idea is to find a way to describe the impact of specific design decisions in a general way.

1.4 Methodical Approach

The methodological approach adopted in this research is designed to systematically explore the concept of spatial interaction for distant displays, identify challenges, develop solutions, and evaluate their effectiveness. In particular, the work on the objectives involves the following two specific research methods centered around the design and prototyping of human-computer interactions. In addition, we employ a third research method of laboratory study to empirically evaluate proposed interactions.

The first is a design space analysis, describing the possibilities within the design of spatial interaction systems and respective interaction techniques. Particularly with smartphone-based systems, new possibilities emerge that we systematically analyze regarding interaction properties that are specific to the technique and commonly used in the literature. We refer to Chapter 2 that summarizes the content of the design space, including smartphone as non-smartphone enabled tracking approaches, hardware devices used, and unimodal as well multimodal interaction techniques.

The second research method is the exploration of the design space through prototyping interaction techniques, user interfaces, and applications for the used system in context. This method is closely coupled with the design space analysis, but is more steered to the technical design of new kinds of interaction artefacts [67]. Artifacts can be basic interaction techniques or whole user interfaces, and this work is also heavily based on application probes [105]. In essence, it is unclear how a new artifact applies to the conventional user interface and how it integrates within a given application context. Prototyping different examples allows us to explore and assess new capabilities quickly, and with it, broaden and refine the design space. Chapters 3, 4 and 5 mainly follow this research method.

The third research method is the laboratory study [63], for testing an interaction technique with users within a controlled environment, often in comparison to a baseline technique. All chapters involve user studies but with different extents. Chapters 3, 4, and 5 provide informal study and focus on qualitative and quantitative user feedback of the developed prototypes and applications. Chapter 3 focuses on a user study of 2D as well as 3D multimodal interaction techniques using hand and touch input in a controlled environment for both qualitative and quantitative performance measures. Chapter 4 focuses on a quantitative user study of user ergonomics and technical tracking range analysis, especially focusing on how users hold the smartphone. Finally, Chapter 4 focuses on a user study of 2D as well as 3D interaction with various multimodal interaction techniques combined out of hand, head, and eye movement with touch in a controlled environment for both qualitative and quantitative performance measures.

1.5 Contributions

We present an inexpensive, accessible, and efficient 6DOF (Degrees Of Freedom) input controller in the form of a smartphone application, which uses a simple and efficient auto-calibration algorithm to adapt the users' control space whenever they change their body position, orientation, or posture. Three studies are conducted to assess the application's usability under various conditions. The findings indicate that users can attain high tracking accuracy in both 2D and 3D interactions with our smartphone-based approach. This level of accuracy is achieved without the need for additional external hardware, such as tracking cameras. Moreover, the application does not force users to execute large hand motions, a key issue with hand-motion-based spatial interaction causing physical hand fatigue. In addition to the studies, we highlight its versatility through demonstrations of controlling various real-world applications.

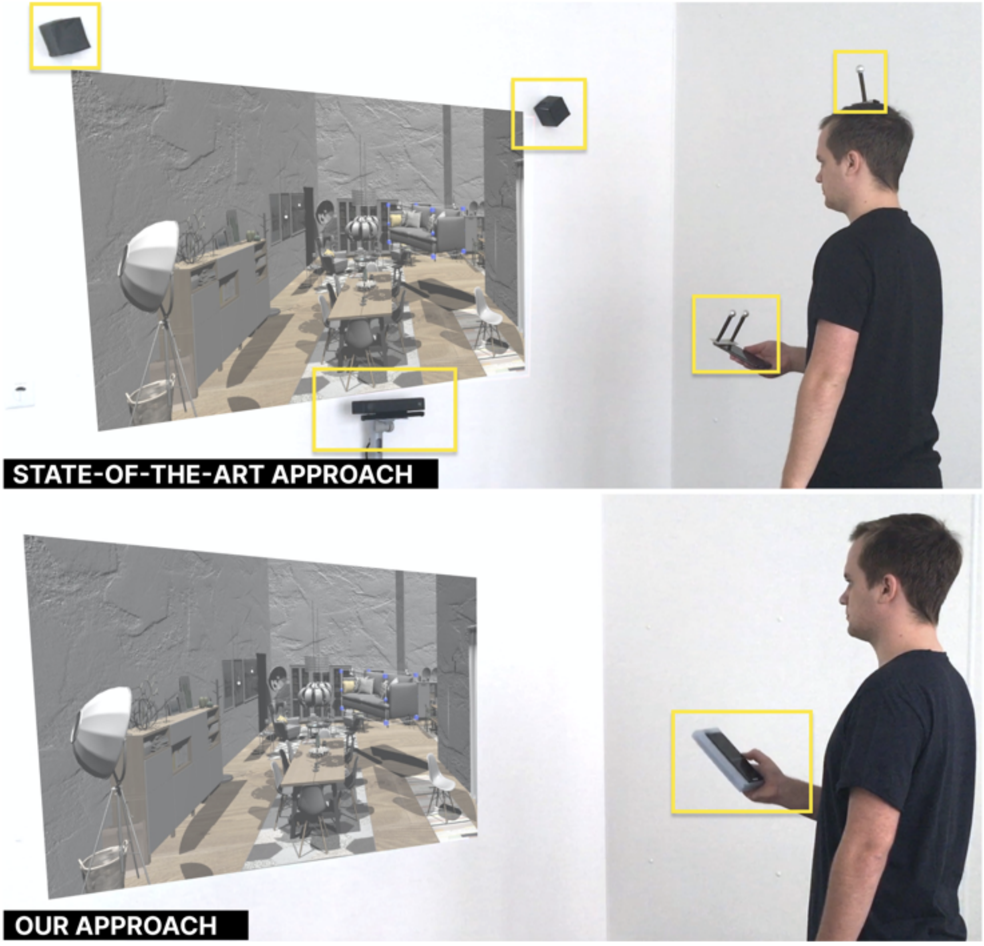


Figure 1.3: Depiction of our innovative smartphone application designed to enable spatial interaction with distant displays. Our solution eliminates the need for placing external tracking devices in the user's environment or the need for users to wear the tracking devices on themselves.

A novel smartphone-based tracking method was introduced that does not require any additional tracking hardware, as seen in Figure 1.3. Utilizing both the front and back cameras of a smartphone simultaneously, this method enables world-scale motion tracking of the user's hand, head, body, and eye-gaze. A user study is presented to not only validate the tracking approach but also provide deeper insights into various multimodal interaction techniques, including refinement techniques for 2D and 3D interactions. Several demonstration applications are showcased to illustrate the practical application of our approach. This tool is set to transform and improve the manner in which researchers develop, distribute (through platforms like app stores), test, and ultimately, how users interact with smartphones and distant displays in their daily lives. Moreover, the approach illustrates that state-of-the-art smartphone technology can facilitate interactions extending well beyond the commonly used touch or device-tilting interactions prevalent today.

Summarizing, the main contributions of this thesis are as follows:

- **Comprehensive Analysis**: The research involved an in-depth analysis of existing spatial interaction systems, identifying their limitations, advantages, and potential areas for improvement. This analysis and the created taxonomy served as the foundation for the subsequent research directions.
- **Design Challenges**: Building on the insights gained from the analysis, three fundamental design challenges were formulated, focusing on enhancing universal design, usability, and flexibility in application for spatial interaction systems, thereby enhancing spatial interaction overall.
- **Concept and Implementation of Smartphone-Based Spatial Interaction**: Leveraging smartphones as versatile tools for spatial interaction, this research repurposed them as trackers and controllers for distant displays. The proposed interaction techniques demonstrated a more universal and portable approach to spatial interaction.
- **Evaluation and Validation**: Rigorous evaluation through user studies helped validate the effectiveness of the proposed interaction techniques. The studies provided valuable quantitative and qualitative feedback, informing iterative refinements across the projects.
- **Multimodal Interaction**: Exploring multimodal interaction techniques, the research combined body motions such as hand and head gestures, touch, as well as eye-gaze for interacting with distant displays. This investigation opened up new possibilities for better usability and standardization of interaction techniques as their adaptation in various applications.
- **Future Research Directions**: The thesis identified several future research directions, building upon the successes and lessons learned from the projects. These directions include exploring other smartphone functionalities for spatial interaction, investigating user interactions with different application scenarios, and developing new interaction techniques based on the research insights.

1.6 Thesis Outline

This thesis is structured in 7 chapters. Each chapter in the thesis, therefore, adds to the exploration of both propositions in the development of a new technique, in the context of a new scenario, or to demonstrate use cases. These are addressed in the main body of this work within Chapters 3, 4, and 5. The insights are discussed in Chapter 6 and concluded in Chapter 7. In the following, we present an overview of the chapters of this thesis:

Chapter 2 Related Work provides an exhaustive review and critical analysis of literature and studies relevant to the thesis's main focus. It delves into the related work of spatial interaction devices and interaction techniques for large interactive distant displays, presenting an analytical taxonomy that outlines the field's current challenges and solutions. Based on the presented taxonomy, in this chapter, three design challenges are synthesized, highlighting the distinctions between the existing state-of-the-art and the contributions presented in this thesis.

Chapter 3 Repurposing the Smartphone into a Spatial Tracker and Controller describes a fully handheld spatial interaction system for distant displays, providing details on a prototype implementation of a motion-tracking smartphone application, its associated interaction techniques, and a user interface for distant displays. The prototype system, named Pocket6, is evaluated in a laboratory study, with a focus on both 2D and 3D interactions that incorporate multimodal inputs containing hand motions and touch inputs. This work enhances understanding of user performance and efficiency when engaging with distant displays through smartphone-based hand motion interactions, all while preserving the user's freedom of motion.

Chapter 4 Extending Smartphones' Spatial Tracking Capabilities by Simultaneous Tracking investigates how the tracking capabilities of the handheld smartphone-based system can be extended to include head and eye tracking. We present the implementation of a prototypical smartphone device consisting of two smartphones used simultaneously. This setup implements and simulates novel tracking capabilities, in particular simultaneous tracking where both front and rear cameras of the smartphones are used concurrently. Based on this prototype, we evaluate the correlation or limitations between smartphone usage ergonomics and the tracking requirements necessary, such as the camera's field of view. The work in this chapter contributes to understanding how we can extend smartphones' interactive capabilities by simultaneous tracking.

Chapter 5 Designing and Evaluating Multimodal Interaction Techniques for 2D and 3D Distant Displays Interaction investigates the enablement of advanced spatial interaction techniques through the use of simultaneous tracking on a single smartphone. We present the implementation of a real (non-prototypical) smartphone application. Following the application's implementation, the chapter outlines the design and implementation of various multimodal interaction techniques, incorporating hand, head, and eye movements, along with touch controls. We evaluate these interaction techniques in 2D as well as 3D interactions. The work extends the knowledge of user performance on spatial interaction techniques placed within the context of everyday distant display interaction.

Chapter 6 Discussion presents a contemplative discussion of the executed research, the implementation on smartphones, and the implications for user interfaces that utilize multimodal interaction techniques, while also situating the work within the framework of previous research. The discussion is guided by a reflection on the initial design challenges and objectives as well as the main research question posed in this thesis.

Chapter 7 Conclusion & Future Work finally presents the concluding remarks of the thesis, limitations and points to future work of smartphone-enabled spatial interaction for distant displays.

Parts of the following Chapter 2 have been published as:

> **Teo Babic**, Harald Reiterer, and Michael Haller. 2022. Understanding and Creating Spatial Interactions with Distant Displays Enabled by Unmodified Off-The-Shelf Smartphones. In: Multimodal Technologies and Interaction. 2022; 6(10):94. DOI: https://doi.org/10.3390/mti6100094

The responsibilities of this publication were as follows: I formulated the research questions, analyzed the existing field of research, designed and created a taxonomy, designed and implemented the research prototype, designed and conducted the user study, analyzed the study data, and wrote the paper. Harald Reiterer and Michael Haller supervised the work.

Supplemental Material

The QR codes below allow you to access the supplemental material by either scanning it using a mobile phone (print) or by clicking on it (digital).

Paper

2

Related Work

Distant displays have become an increasingly prominent feature in modern environments, finding their place in a wide array of settings. These displays are versatile, used in homes for entertainment, in meeting rooms for collaboration, in lecture halls for education, in workplaces for increased productivity, and in public spaces for information dissemination and advertising, fundamentally changing the way information is presented and received.

In this chapter, we firstly unify the fragmented research under the umbrella of distant display and smartphone-based interaction. To present a unified overview and inform future research, we analyzed over 70 papers from this domain, synthesizing the state of the field. We provide an overview of the enabling tracking hardware, input modalities, and interaction techniques of research and commercial systems. We continue by discussing the design challenges of the domain and outlining which interactions should be included in an essential "must-have" bundle of interaction techniques for future distant display scenarios.

2.1 Overview of Approaches to Enable Spatial Interaction with Distant Displays

The technology that supports distant displays has evolved, and with this advancement comes a greater need for effective interaction methods. Engaging with distant displays is not merely about convenience but is a crucial requirement spanning across various applications and contexts. The development and understanding of interaction techniques have emerged as a distinct area of expertise, attracting substantial attention and effort from both researchers and developers aiming to create efficient ways to interact with these displays.

There are many ways to enable interaction with a distant display, and the choice of method may depend on various factors such as the intended use, the audience, the technology available, and the specific requirements of the task. These interaction methods vary widely, from straightforward physical button-based remote controllers to sophisticated body gesture recognition systems.

Figure 2.1: Overview of approaches to enable interaction with a distant display. 1 *Outside-In External Tracking Devices*: Specialized devices in the environment to track user movements. 2 *Inside-Out Dedicated User-Devices*: Dedicated devices, users buy to use the system; these devices themself provide self-contained tracking. 3 *Inside-Out User's Personal Devices*: This approach utilizes embedded technology in everyday devices, emphasizing accessibility and efficiency with a wide range of interactive capabilities.

In the realm of distant display interaction, three predominant tracking methodologies are acknowledged, each embodying distinct paradigms and philosophies (see Figure 2.1). These methodologies, each possess their unique characteristics, benefits, and limitations. They provide the foundational frameworks upon which specific interaction methods can be developed, adapted, and optimized:

- **Outside-In External Tracking Device** approach where interactions are facilitated by specialized tracking devices positioned within the user's environment. These devices might include optical sensors and cameras placed in the room. The precision and adaptability of this approach make it suitable for applications demanding high-fidelity tracking and responsiveness. Common examples are vision-based motion tracking systems, cf. Microsoft Kinect [202, 268] and Figure 2.1 1.
- **Inside-Out Dedicated User-Device** approach where the sensors or tracking mechanisms are located on the device being tracked itself, rather than in the environment around it, as is the case with outside-in external tracking devices. These can include cameras or other sensors that are part of the controller device, often allowing it to observe its surroundings and determine its position and orientation within a space. For non-spatial interfaces, it could simply be a button or a touch-based remote controller. In spatial interaction, this method usually depends on devices like dedicated motion-sensitive handheld controllers (see Figure 2.1 2) or wearables [14, 152], rather than personal devices like smartphones. The inside-out tracking method provides self-contained tracking, making the device adaptable to various environments without the need for external sensors or markers.
- **Inside-Out User's Personal Device** approach that relies solely on the embedded technology within everyday personal devices, such as smartphones, smartwatches, rings, or smart-glasses (see Figure 2.1 3). Utilizing existing sensors like accelerometers, gyroscopes, or touch interfaces, these inside-out devices can recognize and interpret user interactions without the need for external tracking equipment [171, 200]. The personal devices approach emphasizes accessibility and efficiency, leveraging commonly owned devices to achieve a wide range of interactive capabilities.

Next, we will dive deeper into the characteristics inherent to all three approaches.

2.1.1 Spatial Interaction System Based on Dedicated Devices

Numerous researchers have examined the use of dedicated tracking devices like Microsoft Kinect, OptiTrack motion tracking systems, or compact desktop trackers like Leap Motion, assessing their compatibility and functionality with distant displays. They investigated the application of movements from different body parts for interaction, encompassing body position and orientation, hand motion, and concurrent movements from multiple users.

A review of extensive related work on outside-in trackers, such as vision-based lighthouse trackers, reveals that these predominantly prototypical in-lab systems present numerous limitations. These include difficulty in setup and replication, high costs, substantial space requirements, lack of mobility, calibration necessities, additional computational unit requirements, power consumption, limited tracking range and field-of-view, and the need for operation in controlled environments with a predetermined number of users moving in specific ways to avoid camera occlusion. On a technical level, there is a need in the field of spatial interaction research for more pragmatic solutions that facilitate testing and refinement in settings beyond the laboratory, thereby supporting broader use and deployments in real-world, in-the-wild scenarios. While academics and industry professionals have started addressing these infrastructure challenges, there has not been a concentrated effort towards reducing hardware setup complexities to facilitate an immediate, out-of-the-box experience [39].

A large amount of work in this domain has focused on hand gestures due to the expressive potential they offer to users. The growing interest in mid-air hand gestures stems from their "come as you are" principle to interact with distant displays. By translating physical movements in three-dimensional space, these hand gestures allow users to control and manipulate digital content in a way that mirrors real-world interactions with real objects. Advances in precise sensing technologies and complex algorithms have facilitated the translation of these gestures into distinct commands, effectively connecting physical actions with digital responses. Various gestural interaction techniques, such as semaphoric, pointing, and direct manipulation, have been the subject of extensive investigation by researchers [2]. The field of natural user interfaces faces numerous challenges associated with the use of hand gesture input, presenting a complex set of problems to solve [27, 185, 284]. Challenges include physical hand fatigue [98, 217], difficulty in reaching specific spatial areas [144], acceptance of the interaction method [104], the necessity for users to memorize particular gesture sets [179], and the Midas problem, where unintentional gestures may trigger undesired actions [144, 284]. Despite the improvement in sensor capabilities, occlusion issues due to the physical distance between the sensor and the user will remain, necessitating users to execute large and clear hand or arm gestures clearly visible in front of their bodies [152]. Such larger gestures are not only more physically exhausting but also socially awkward and challenging to execute in confined spaces [98, 111]. Researchers propose that addressing these challenges requires a comprehensive approach to designing new user interfaces, focusing on factors like user skill, learning ability, and investment of time as core components. These elements are particularly vital in the context of post-WIMP interfaces, where merely having an effective recognition system, a memorable set of gestures, or accurate visualization individually is not sufficient to guarantee a successful product. Instead, the whole “package” needs to be thoughtfully designed [36, 284].

As previously mentioned, numerous approaches focus on creating interaction techniques that do not necessitate the use of any dedicated devices, adhering to a "come as you are" philosophy. In contrast, other researchers opt for the method of employing specialized (non-personal) controllers to enable spatial interactions with distant displays. Using specialized motion trackers or handheld controllers to interact with distant displays has, however, considerable disadvantages which are very similar to dedicated outside-in tracking systems (mentioned previously) [118]. The expense of these devices poses a significant obstacle, especially for smaller entities or individual users for whom the cost may be prohibitive. The devices’ lack of mobility presents another hurdle, as their need for specific setups or calibrations often confines their use to particular environments. Such limitations can limit the flexibility and spontaneity in interactions, possibly inhibiting creative or collaborative endeavors. The devices’ power requirements add to the complexity of using these devices. The need for regular charging can be inconvenient, and the potential for power failure during crucial interactions is a risk that must be managed. This additional layer of planning and consideration might discourage some users from adopting this technology. Furthermore, the restricted tracking range of some devices can adversely affect the user experience. If the devices are unable to accurately track movements beyond a specific range, this can lead to issues and user frustration. Such limitations might restrict the kinds of applications and experiences effectively delivered through these technologies. Despite these challenges, dedicated handheld controllers and motion trackers have played a crucial role in advancing spatial interaction with distant displays. Nonetheless, addressing these drawbacks is needed, or there must be an exploration of alternative solutions to enhance the accessibility and versatility of this technology in diverse contexts.

To avoid the need for instrumenting the environment with cameras, as seen with external tracking devices, or the purchase of dedicated controllers, employing personal devices like smartphones emerges as the third alternative for enabling spatial interaction with distant displays. This approach is intriguing to researchers due to the ubiquitous presence of smartphones, with billions consistently carrying them around at all times [33, 234].

2.1.2 Spatial Interaction System Based on Personal Devices

As we navigate the journey of enabling spatial interaction for distant displays, one ubiquitous device stands out with immense potential – the smartphone. Today, the smartphone has become an extension of ourselves, a device we carry everywhere, and use for a myriad of purposes. Its widespread use, combined with inherent technical features and capabilities, makes the smartphone a promising candidate for enabling spatial interaction. The rationale and motivation behind repurposing smartphones, originates from acknowledging the smartphone's inherent capabilities. With an array of sensors, high-resolution touch screens, and vast computing power, smartphones can run complex applications recognizing and interpreting complex user actions and gestures. Such features can be leveraged to facilitate spatial interaction with distant displays, transforming the smartphone into an effective spatial controller and tracker.

The following chapter will examine the concept of repurposing smartphones as proposed by related researchers in more detail. The analysis will begin with the creation of a taxonomy to comprehend the range of smartphone functionalities relevant to spatial interaction.

2.2 Analysis of Approaches Relying on Personal Devices to Enable Spatial Interaction with Distant Displays

In pursuit of a novel approach to user-motion tracking that accommodates a broad range of users, we asked the following question: Which input and output devices are universally accessible for everyone? With this in mind, we focused our research on enabling spatial inputs for a distant display by an off-the-shelf device which is predominantly accessed and utilized by the majority of individuals — the smartphone.

In the subsequent section, we introduce a taxonomy that outlines known related smartphone-based approaches and highlights the smartphone functionalities relevant to spatial interaction. The goal of our taxonomy is to contribute a comprehensive analysis of the design space where smartphones, with or without additional tracking devices, are used to drive the interaction with a distant display. For both new and experienced researchers, we provide an overview of tracking technology, input modalities, and interaction techniques, spanning from 2D to 3D distant display interactions. Additionally, we delve into the tradeoffs and implications of utilizing tracking hardware, exploring its influence on the creative process of crafting interaction designs and the resulting interaction techniques.

We build upon other related taxonomies, all with their own specialized focus covering:

- Cross-device interaction [39]
- Sensing techniques for mobile devices [100]
- Eye-gaze interaction [125]
- Augmented reality enabled mobile devices [223, 276]
- Proxemic relationships [161]
- Spatial input by handhelds [71, 99, 261]

2.2.1 Designing the Taxonomy

To assemble the main corpus of relevant publications, a systematic search within the field of HCI was conducted, encompassing an analysis of 405 papers spanning the last 50 years. The keywords we were primarily interested, were: *phone*, *smartphone*, *mobile*, *handheld*, *controller*, *spatial*, *vertical*, *spatially-aware*, *cross-device*, *distant*, *public*, *situated*, *remote*, *large*, *pervasive*, *wall*, *display*, *interaction*, *interfaces*, as well as their respective acronyms. Papers selected for inclusion in our corpus were required to address spatial interaction tasks or techniques and tracking technology applicable to individuals and/or devices. By looking at references within our corpus as well as using our own expertise, we identified additional articles (which is a common strategy for survey and taxonomy papers [39]). Within the initial corpus, 405 papers were tagged based on their utilization of a *distant display* or a *smartphone* in interactive setups, resulting in three subsets:

- 73 papers using smartphones alongside distant displays, optionally with additional tracking devices.
- 157 papers using distant displays with input devices other than smartphones (e.g. wands, smartwatches, lighthouse trackers).
- 175 papers exclusively using off-the-shelf smartphones (e.g. on-phone graphical user interfaces) or using them in combination with devices such as eye-trackers, pens, head-mounted displays, or other mobile gadgets.

Based on our specific focus, we included 73 papers using a smartphone *and* a distant display (concurrently) into our taxonomy. Naturally, we also considered older mobile devices (without touchscreens) and tablets which were used with a distant display. In the next step, we tagged all these papers based on *tracking hardware*, *input modalities*, and *interaction techniques* and created overview figures, which construction we will explain next.

Figure 2.2: The structure of the taxonomy overview figures. A *Input Tracking* showcases the primary sources of interaction, either through device manipulation or body movements. B *Hardware Setup Type* details the configurations of tracking devices, distinguishing between smartphone-only and hybrid setups. C *Input Modalities* breaks down the specific input modalities or motion axes used for interaction. D *Interaction Techniques* elucidates the mapping of input modalities to spatial interactions.

The taxonomy is presented and described in the following Section 2.2.2. The arrangement within and across this taxonomy and the created figures conforms to a specific structure, explained in the following and accompanied by the Figure 2.2:

- **Input Tracking** (Figure 2.2 A): This primary category of the taxonomy structure explains what the hardware trackers primarily track. We present two options: *device-tracking inputs*, where interactions arise from manipulating the smartphone itself, either by moving it spatially or using its touchscreen, and *body-tracking inputs*, where user body movements generate inputs, such as when a user moves its head.
- **Hardware Setup Type** (Figure 2.2 B): This category in the taxonomy structure details the devices utilized for tracking. Two configurations are highlighted: *inside-out*, where smartphone-only is used, and *hybrid*, where additional tracking devices accompany the smartphone.
- **Input Modalities** (Figure 2.2 C): This segment of the taxonomy outlines the specific modality or, in the context of our spatial interaction, the precise smartphone motion axis or user body parts involved in the interaction. The options for *device-tracking inputs* include: *smartphone touch* for inputs created via the touchscreen, *smartphone tilting* for inputs based on smart-

phone's orientation changes, *smartphone translation* for input created from position changes, *smartphone pose* for inputs created by both orientation and position changes, and *smartphone user interface* for inputs created through the device's graphical user interface. The options for *body-tracking inputs* include: *traveling or walking* for inputs created by orientation and position changes of the user's body position, *head motion* for inputs created through tracking the orientation and position of the users head, and *eye motion or gaze* for inputs based on tracking the user's gaze direction.

- **Interaction Techniques** (Figure 2.2 D): This final category explains how exactly the previously mentioned input modalities were mapped to the spatial interaction techniques. For instance, a vertical x-axis finger swipe on the touchscreen corresponds to the z-axis rotation of a 3D object on a distant display [12].

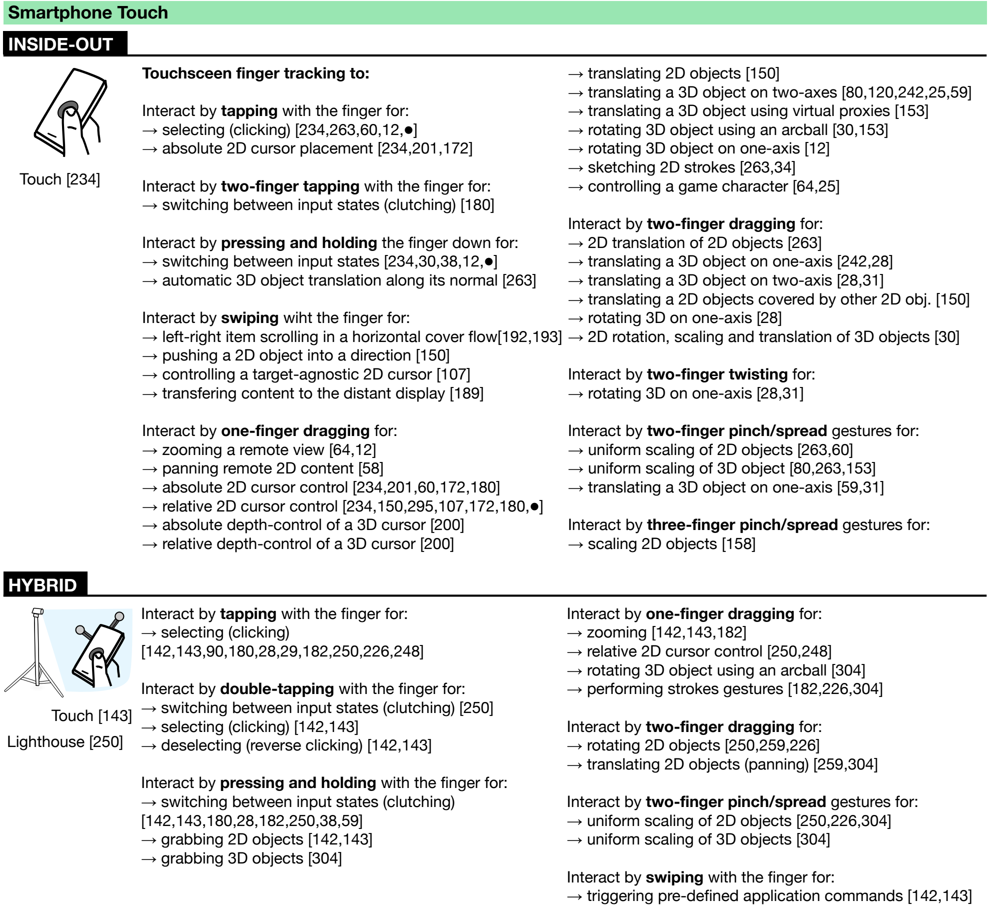
2.2.2 Taxonomy

This section presents a concise overview of the taxonomy, laying the foundation for an in-depth analysis and discussion of the key findings, which will be elaborated upon in the remaining sections of this chapter. We start by presenting the category of *device-tracking inputs*, followed by the *body-tracking inputs* category.

Device-Tracking Inputs

The *device-tracking inputs* category in the taxonomy focuses on interaction techniques that utilize the tracking capabilities of the smartphone device itself for spatial interaction. This includes methods demonstrating how utilizing the smartphone's touchscreen or the motion of the device can be translated into interactive commands.

Smartphone Touch This section and Figure 2.3 outline interaction methods that utilize the smartphone's touchscreen. These include single and multi-touch gestures such as tapping, swiping, pinching, and spreading. Each gesture is associated with specific actions such as selecting, cursor placement, scrolling, transferring content, and controlling 3D objects. The diversity in touch-based interactions highlights the smartphone's versatility in managing both 2D and 3D objects in spatial interaction contexts. While foundational touch interactions as tapping, double-tapping, or swipe gestures are consistent across both *inside-out* and *hybrid* systems, a direct comparison between these two categories does not reveal a significant difference in their overall system capabilities, regardless of the use of additional tracking hardware. The detailed listing of interaction techniques primarily reveals the exact touch interactions selected by researchers and designers in each setup.



Smartphone Touch

INSIDE-OUT

Touch [234]

Touchsceen finger tracking to:

Interact by **tapping** with the finger for:
→ selecting (clicking) [234,263,60,12,●]
→ absolute 2D cursor placement [234,201,172]

Interact by **two-finger tapping** with the finger for:
→ switching between input states (clutching) [180]

Interact by **pressing and holding** the finger down for:
→ switching between input states [234,30,38,12,●]
→ automatic 3D object translation along its normal [263]

Interact by **swiping** wiht the finger for:
→ left-right item scrolling in a horizontal cover flow[192,193]
→ pushing a 2D object into a direction [150]
→ controlling a target-agnostic 2D cursor [107]
→ transfering content to the distant display [189]

Interact by **one-finger dragging** for:
→ zooming a remote view [64,12]
→ panning remote 2D content [58]
→ absolute 2D cursor control [234,201,60,172,180]
→ relative 2D cursor control [234,150,295,107,172,180,●]
→ absolute depth-control of a 3D cursor [200]
→ relative depth-control of a 3D cursor [200]
→ translating 2D objects [150]
→ translating a 3D object on two-axes [80,120,242,25,59]
→ translating a 3D object using virtual proxies [153]
→ rotating 3D object using an arcball [30,153]
→ rotating 3D object on one-axis [12]
→ sketching 2D strokes [263,34]
→ controlling a game character [64,25]

Interact by **two-finger dragging** for:
→ 2D translation of 2D objects [263]
→ translating a 3D object on one-axis [242,28]
→ translating a 3D object on two-axis [28,31]
→ translating a 2D objects covered by other 2D obj. [150]
→ rotating 3D on one-axis [28]
→ 2D rotation, scaling and translation of 3D objects [30]

Interact by **two-finger twisting** for:
→ rotating 3D on one-axis [28,31]

Interact by **two-finger pinch/spread** gestures for:
→ uniform scaling of 2D objects [263,60]
→ uniform scaling of 3D object [80,263,153]
→ translating a 3D object on one-axis [59,31]

Interact by **three-finger pinch/spread** gestures for:
→ scaling 2D objects [158]

HYBRID

Touch [143]
Lighthouse [250]

Interact by **tapping** with the finger for:
→ selecting (clicking)
[142,143,90,180,28,29,182,250,226,248]

Interact by **double-tapping** with the finger for:
→ switching between input states (clutching) [250]
→ selecting (clicking) [142,143]
→ deselecting (reverse clicking) [142,143]

Interact by **pressing and holding** with the finger for:
→ switching between input states (clutching)
[142,143,180,28,182,250,38,59]
→ grabbing 2D objects [142,143]
→ grabbing 3D objects [304]

Interact by **one-finger dragging** for:
→ zooming [142,143,182]
→ relative 2D cursor control [250,248]
→ rotating 3D object using an arcball [304]
→ performing strokes gestures [182,226,304]

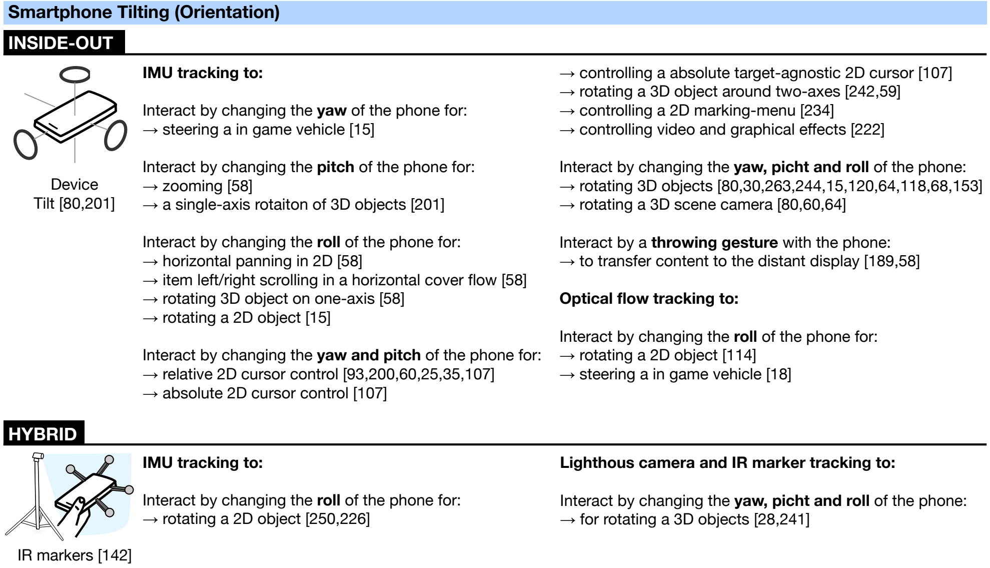
Interact by **two-finger dragging** for:
→ rotating 2D objects [250,259,226]
→ translating 2D objects (panning) [259,304]

Interact by **two-finger pinch/spread** gestures for:
→ uniform scaling of 2D objects [250,226,304]
→ uniform scaling of 3D objects [304]

Interact by **swiping** with the finger for:
→ triggering pre-defined application commands [142,143]

Figure 2.3: Figure showing the range of *smartphone touch* interaction techniques for *inside-out* smartphone only and *hybrid* setups that use additional hardware. Indicated with ● are tracking abilities enabled by the system proposed in this thesis.

Smartphone Tilting (Orientation) In *inside-out* systems, tilting-based interactions are enabled by the smartphone's inertial measurement unit (IMU), as shown in Figure 2.4. The overview details how tilting the phone along different axes (yaw, pitch, roll) enables a range of actions like cursor pointing, zooming, panning, rotating objects, or steering vehicles in games. The tilt-based interactions showcase how to engage with digital content, merging physical movements with virtual responses. The addition of external devices, such as a lighthouse camera system or IR marker tracking, enriches device orientation-based tilting interaction, enabling more precise control and detection of the phone's roll, pitch, and yaw changes.



Smartphone Tilting (Orientation)

INSIDE-OUT

Device Tilt [80,201]

IMU tracking to:

Interact by changing the **yaw** of the phone for:
→ steering a in game vehicle [15]

Interact by changing the **pitch** of the phone for:
→ zooming [58]
→ a single-axis rotaiton of 3D objects [201]

Interact by changing the **roll** of the phone for:
→ horizontal panning in 2D [58]
→ item left/right scrolling in a horizontal cover flow [58]
→ rotating 3D object on one-axis [58]
→ rotating a 2D object [15]

Interact by changing the **yaw and pitch** of the phone for:
→ relative 2D cursor control [93,200,60,25,35,107]
→ absolute 2D cursor control [107]
→ controlling a absolute target-agnostic 2D cursor [107]
→ rotating a 3D object around two-axes [242,59]
→ controlling a 2D marking-menu [234]
→ controlling video and graphical effects [222]

Interact by changing the **yaw, picht and roll** of the phone:
→ rotating 3D objects [80,30,263,244,15,120,64,118,68,153]
→ rotating a 3D scene camera [80,60,64]

Interact by a **throwing gesture** with the phone:
→ to transfer content to the distant display [189,58]

Optical flow tracking to:

Interact by changing the **roll** of the phone for:
→ rotating a 2D object [114]
→ steering a in game vehicle [18]

HYBRID

IR markers [142]

IMU tracking to:

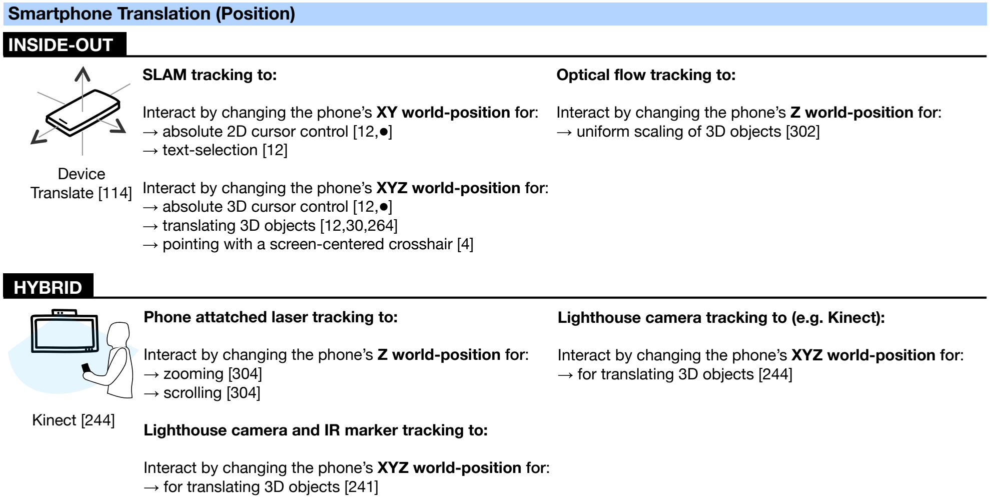
Interact by changing the **roll** of the phone for:
→ rotating a 2D object [250,226]

Lighthous camera and IR marker tracking to:

Interact by changing the **yaw, picht and roll** of the phone:
→ for rotating a 3D objects [28,241]

Figure 2.4: Figure showing the range of *smartphone tilting* interaction techniques for *inside-out* smartphone only and *hybrid* setups that use additional hardware.

Smartphone Translation (Position) In *inside-out* systems, the smartphone's ability to track its position in world space is utilized by using technologies such as SLAM (Simultaneous Localization and Mapping), as seen in Figure 2.5. This tracking capability allows for precise control of cursors and objects in both 2D and 3D environments. Enhancements by phone-attached laser tracking or the Kinect tracking system allowed in *hybrid* systems, even before the utilization of SLAM to smartphones, translation tracking of the phone in the *XYZ* world space. This enabled advancements in spatial interaction such as zooming, scrolling, and 3D object manipulation, and more detailed interaction possibilities in both 2D and 3D user interfaces.



Smartphone Translation (Position)

INSIDE-OUT

Device Translate [114]

SLAM tracking to:

Interact by changing the phone's **XY world-position** for:
→ absolute 2D cursor control [12,●]
→ text-selection [12]

Interact by changing the phone's **XYZ world-position** for:
→ absolute 3D cursor control [12,●]
→ translating 3D objects [12,30,264]
→ pointing with a screen-centered crosshair [4]

Optical flow tracking to:

Interact by changing the phone's **Z world-position** for:
→ uniform scaling of 3D objects [302]

HYBRID

Kinect [244]

Phone attatched laser tracking to:

Interact by changing the phone's **Z world-position** for:
→ zooming [304]
→ scrolling [304]

Lighthouse camera and IR marker tracking to:

Interact by changing the phone's **XYZ world-position** for:
→ for translating 3D objects [241]

Lighthouse camera tracking to (e.g. Kinect):

Interact by changing the phone's **XYZ world-position** for:
→ for translating 3D objects [244]

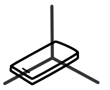
Figure 2.5: Figure showing the range of *smartphone translation* interaction techniques for *inside-out* smartphone only and *hybrid* setups with additional hardware. Indicated with ● are tracking abilities enabled by the system proposed in this thesis.

Smartphone Pose (Position + Orientation) Pose tracking combines the smartphone's position and orientation tracking capabilities. Consequently, the *smartphone pose* is being enabled by the same tracking technology as the *smartphone tilting* and *smartphone translation* tracking. In *inside-out* systems, pose tracking is primarily enabled with SLAM technology, and in *hybrid* systems with lighthouse, laser, or IR marker tracking, as shown in Figure 2.6. The overview shows how moving the phone in world-space can be used for controlling 2D and 3D cursors or facilitating interactions that mimic natural hand gestures and movements.

Smartphone User Interface This final interaction method in the *device-tracking inputs* category emphasizes inputs created through the phone's graphical user interface, such as using buttons, sliders, and lists to control remote interfaces, games, or manipulate 3D objects, as depicted in 2.7. It also introduces concepts such as peephole [55] and distributed user interfaces [69], underscoring the smartphone's role as a bridge between the user and the digital world. In *hybrid* setups, the interactions through the phone's user interface remain consistent with the *smartphone-only* category, underscoring the general applicability of the smartphone's native graphical user interface capabilities.

Smartphone Pose (Postion + Orientation)

INSIDE-OUT

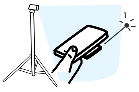


Device Pose [12]

SLAM tracking to:

Interact by **moving the phone in world-space** for:
→ absolute 2D cursor control [●]
→ absolute 3D cursor control (virtual-hand) [30,12,●]
→ absolute 3D cursor control (depth ray-casting) [●]

HYBRID



Laser [226]

Lighthouse camera and phone attached laser tracking to:

Interact by **moving the phone in world-space** for:
→ absolute 2D cursor control (ray-casting) [304]
→ absolute 2D cursor control (ray-casting) [226,177,194]

Lighthouse camera and IR marker tracking to:

Interact by **moving the phone in world-space** for:
→ absolute 2D cursor control (ray-casting) [142,143,90,130,292,182,226]
→ absolute 3D cursor control (virtual-hand) [28,29]
→ translating and rotating 3D objects [28]

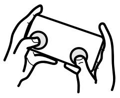
Lighthouse camera tracking to (e.g. Kinect):

Interact by **moving the phone in world-space** for:
→ absolute 2D cursor control (ray-casting) [38]

Figure 2.6: Figure showing the range of *smartphone pose* interaction techniques for *inside-out* smartphone only and *hybrid* setups that use additional hardware. Indicated with ● are tracking abilities enabled by the system proposed in this thesis.

Smartphone User Interface

INSIDE-OUT



UI [15,33]

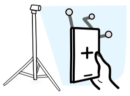
Phone's graphical user interface to:

Interact by **UI elements** (e.g. buttons, sliders, list):
→ for controlling a remote user interface[193,26,15,295]
→ for controling a game character [193,262]
→ for controling translating 3D objects [29]

Interact by a **peephole user interface** to:
→ project phone-touch to the distant display [33,18]

Interact by a **distributed user interface** to:
→ intput data [84,4,122]
→ exchange content [13]

HYBRID



Peephole [205]

Phone's graphical user interface to:

Interact by **UI elements** (e.g. buttons, sliders, list):
→ for controlling a remote user interface [141,130,241]

Interact by a **peephole user interface**:
→ for controlling a remote user interface [141,130,205]
→ to project phone-touch to the disntat display [90]

Interact by a **distributed user interface** to:
→ intput data [130,241]

Figure 2.7: Figure showing the range of *smartphone user interface* interaction techniques for *inside-out* smartphone only and *hybrid* setups with additional hardware.

Body-Tracking Inputs

The *body-tracking inputs* category in our taxonomy concentrates on interaction techniques that leverage the user's body movements for spatial interactions. This encompasses methods where gestures of individual body parts and or the full body posture, tracked via smartphone sensors or additional hardware, translate into inputs for spatial interaction with distant displays.

Traveling / Walking User travel tracking describes spatial interaction through whole-body movement. For tracking *inside-out* systems leverage the smartphone's simultaneous face- and world-tracking, while *hybrid* systems incorporate additional lighthouse camera and IR marker tracking hardware. Both approaches focus on the user's body position and orientation, enabling interactive experiences based on real-world movement or stance of the user in front of a distant display, as shown in Figure 2.8. Furthermore, several examples demonstrate how body tracking not only enables input methods but also enhances output, such as interactions that adjust a 3D camera's viewpoint using the user's body position and leaning [59].

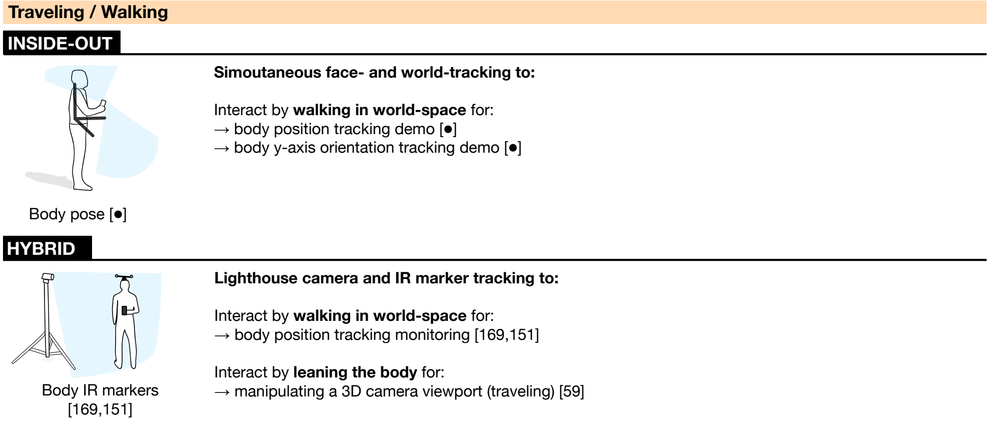
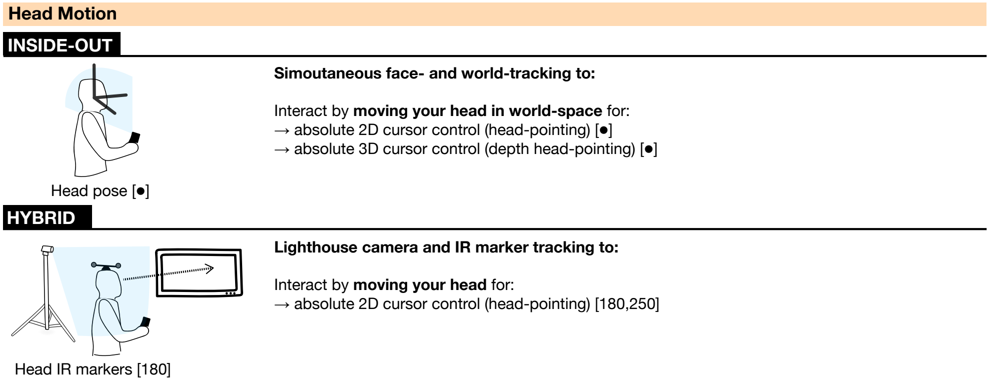


Figure 2.8: Figure showing the range of *traveling / walking* interaction techniques for *inside-out* smartphone only and *hybrid* setups with additional hardware. Indicated with ● are tracking abilities enabled by the system proposed in this thesis.

Head Motion In the *inside-out* systems, head pose tracking can be utilized by using simultaneous face- and world-tracking. This enables absolute 2D and 3D cursor control through head-pointing techniques, where users can interact by moving their head, as depicted in Figure 2.9. Meanwhile, the *hybrid* setups again often incorporate lighthouse camera and IR marker tracking, offering similar head-pointing capabilities for absolute 2D cursor control. These methods showcase the use of head movements for precise cursor manipulation in both 2D and 3D user interfaces presented in distant displays.



Head Motion

INSIDE-OUT

Head pose [●]

Simoutaneous face- and world-tracking to:

Interact by **moving your head in world-space** for:
→ absolute 2D cursor control (head-pointing) [●]
→ absolute 3D cursor control (depth head-pointing) [●]

HYBRID

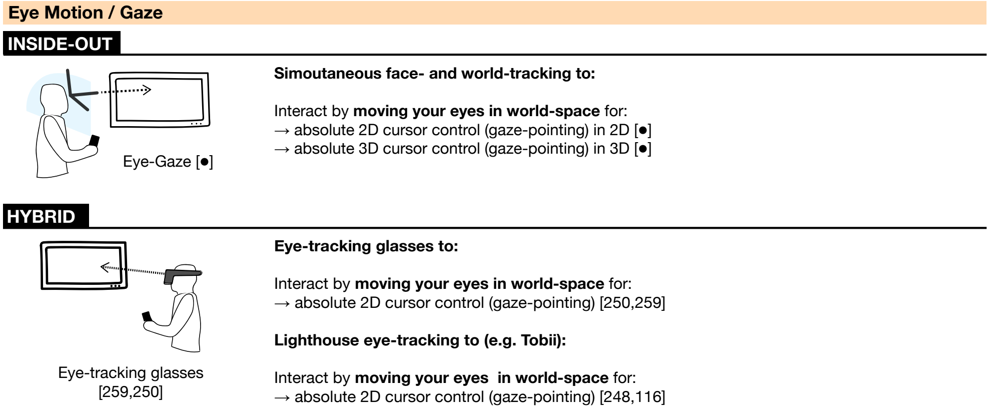
Head IR markers [180]

Lighthouse camera and IR marker tracking to:

Interact by **moving your head** for:
→ absolute 2D cursor control (head-pointing) [180,250]

Figure 2.9: Figure showing the range of *head motion* interaction techniques for *inside-out* smartphone only and *hybrid* setups with additional hardware. Indicated with ● are tracking abilities enabled by the system proposed in this thesis.

Eye Motion / Gaze Similar to head-motion tracking in *inside-out* systems, simultaneous face- and world-tracking technologies can be used to facilitate eye-gaze tracking. This enables interaction by moving the eyes in world-space, offering a direct gaze-pointing cursor control in both 2D and 3D distant displays user interfaces. In the *hybrid* systems, additional hardware like eye-tracking glasses and specialized devices (e.g., Tobii lighthouse eye-tracking) are incorporated to enable gaze-based interaction. These systems also focus on absolute 2D cursor control via gaze-pointing.



Eye Motion / Gaze

INSIDE-OUT

Eye-Gaze [●]

Simoutaneous face- and world-tracking to:

Interact by **moving your eyes in world-space** for:
→ absolute 2D cursor control (gaze-pointing) in 2D [●]
→ absolute 3D cursor control (gaze-pointing) in 3D [●]

HYBRID

Eye-tracking glasses
[259,250]

Eye-tracking glasses to:

Interact by **moving your eyes in world-space** for:
→ absolute 2D cursor control (gaze-pointing) [250,259]

Lighthouse eye-tracking to (e.g. Tobii):

Interact by **moving your eyes in world-space** for:
→ absolute 2D cursor control (gaze-pointing) [248,116]

Figure 2.10: Figure showing the range of *eye motion / gaze* interaction techniques for *inside-out* smartphone only and *hybrid* setups with additional hardware. Indicated with ● are tracking abilities enabled by the system proposed in this thesis.

In the following sections, we will delve deeper into each category, discussing their characteristics, comparing them with one another, and highlighting their relevance for this thesis and spatial interaction with distant displays.

2.2.3 Types of Tracking Systems

By analyzing related papers, we agree with Brudy et al. [39], that the tracking systems for enabling interactions with distant displays can be broadly categorized into into two types: *inside-out* and *outside-in*, regardless the specific device used for tracking. Upon a more detailed review of relevant related work, we identified that these categories could be further refined to more accurately categorize interaction enabled by smartphone devices with distant displays. This resulted in the definition of the following two primary tracking types for smartphone-based distant display tracking systems:

- **Inside-out Smartphone Only** interactions, enabled solely by using the technology, embedded in off-the-shelf smartphones. This category investigated approaches using only the smartphone alone to enable touch-based [181, 250], device tilting [201, 234] or swinging [58, 188] based 2D [35, 264] as well as 3D [95, 201] interactions with distant displays.
- **Hybrid** interactions, enabled by using inside-out tracking capabilities of smartphones, which get further enhanced by additional tracking devices (i.e. tracking devices attached to the smartphone, worn by the user, or placed in the room). The second group, used the smartphone together with additional external motion trackers, mainly to enable tracking of the user's head [170, 260] or body [113, 269]. These works aimed for more advanced input methods, showing how smartphone touch or device-tilting 2D and 3D interactions can be further enriched by head-pointing, body-motion or egocentric inputs enabled by the additional device.

Although the *hybrid* approaches diminished the major advantage of *smartphone-only* approaches, as its large user base and effortless use which does not require any additional hardware, they demonstrate various use cases for distant displays where recognizing the user's body position or gaze direction significantly enhances the interaction. Hereby, we believe that a smartphone-only solution where users could just download an app and get access to touch, hand, head, and body motion-based input capabilities, which would work out-of-the-box without additional hardware and setup, would be highly beneficial.

EXTRAS

Distant display + X

Vision-based lighthouse tracking for interactions based on:
bare-handed [166, 109, 158, 160, 154, 47, 160, 268, 184, 269, 52]
finger-gestures [66]
body pose [113, 269, 11, 3, 109, 166, 136]
body outline [230, 229, 278]
body skeleton [169]
eye-gaze [47, 301, 158, 127, 254, 199, 238]
head-based [184, 196, 257]
tangible objects [52, 30, 99, 123, 24, 297]

Drone-based tracking [128]

Wearable devices:
wearable eye-tracking glasses for gaze interaction [140, 231, 298, 220]
smartwatches [234, 101, 233, 117, 102]
shoulder-worn system [88]
bands[152, 86, 293]
textiles [191, 221]
shoes [14]
rings [291]

Handheld devices:
wands [146, 113, 23, 171, 115, 286]
handheld remote controllers [17, 236]
laser pointers [132, 270, 177, 186]
pens [72, 129]
joysticks [236]

On-Display inputs:
Large touchscreens [283, 108, 150, 294]
phones in physical contact with the distant display [148, 237]
speech [32]

3D user interfaces:
3D user interfaces [249, 296]
fish-tank user interfaces [146, 224, 281, 209, 282]
caves [57]

Smartphone only | Smartphone + X

Interactions:
touch [145, 5, 82, 53, 255]
gaze [197, 267, 126, 300, 134, 159, 213]
head [147, 266, 168, 51]
phone's casing [228, 227, 78]
foot [155]
hardware buttons [289]
barehanded gestures [13, 8, 92, 139]
finger in mid-air [258]
head-mounted displays [303, 50, 255, 174, 40]
cross-device interactions [207, 206, 162]
smartwatch [48]
pen [272, 75, 273, 198, 211]

User interfaces:
handheld augmented-reality [272, 55, 275, 276, 96, 94, 175, 274, 168]
spatially-aware [70, 212, 280, 1, 216, 139, 256]
peephole [73, 42, 41, 91, 246, 190, 49, 215, 290]
magic lens [173, 21, 22, 245, 97]
phone-projector [121, 287, 285]

Phone Keys and Joysticks (legacy)

Keys [87]

Phone hardware keys and joysticks to:

Interact by **pressing** a key for:
→ selecting ("click") [80, 120]

Interact by **pressing and holding a key** for:
→ switching the input state (clutching) [87, 68]

Interact **by tilting the joystick** for:
→ relative 2D cursor control [177, 120]

Figure 2.11: A list of domains that are out of the scope of this work could, however, directly benefit from our approach to initially enable or enhance their existing spatial interactions. The overview is structured based on three categories: First, *Distant Display + X* present approaches where spatial interaction with a distant display was enabled without incorporating a smartphone into the setup, by using another device *X*. Second, *Smartphone Only and Smartphone + X* presents approaches where a smartphone was used to enable spatial interaction with other devices *X*, outside of a distant display context. Finally, third, *Smartphone Keys and Joysticks (legacy)* presents approaches where the smartphone's hardware keys or joysticks enable spatial interaction. We label this as a legacy approach, given the rarity of current smartphones featuring hardware keys or joysticks.

General Bias Associated with Different Tracking System Types

Upon reviewing relevant research, we noticed that researchers commonly conclude that outside-in tracking provides high-fidelity data, while inside-out tracking technology, despite being more lightweight, delivers information of lower fidelity. Therefore, the primary challenge for new inside-out smartphone-based approaches lies in reliably tracking the smartphone in 3D world-space and precisely incorporating the user's motion into the sensing mechanism, as noted by Brudy et al. [39]. Nevertheless, it is noteworthy that only very few related projects on inside-out systems have addressed this particular issue to date. Consequently, we believe that this sentiment, suggesting that simpler inside-out systems cannot achieve the performance of complex outside-in systems, will hold unless an inside-out system will not be able to achieve the same tracking precision and recognize as many input modalities as its outside-in counterparts.

Inform Related Domains

While our research focuses on *Inside-out Smartphone Only* and *Hybrid* categories, we also acknowledge the relevance of other categories beyond the scope of this thesis. These categories, though not directly addressed in our work, could potentially benefit from our taxonomy and the presented tracking approach. Figure 2.11 shows in which other scenarios beyond the distant display context, the smartphone-based user-motion tracking could be used to initially enable or enhance existing spatial interactions. We outline systems where a non-smartphone device facilitated interaction with a distant display and systems where a smartphone enabled interaction with another, not distant display, device.

- **Distant Display + X:** Present approaches where spatial interaction with a distant display was enabled by a non-smartphone device *X*, such as enabling a spatial interaction with the utility of a wearable ring [299], or drone-based tracking [128].
- **Smartphone Only and Smartphone + X:** Presents approach where a smartphone was used to enable spatial interaction with another device or context *X*, outside of a distant display context. This includes approaches that enable spatial interactions with the smartphone itself (*Smartphone Only*) or scenarios where the smartphone was paired with another device for spatial interactions between devices, such as between a smartphone and a

smartwatch [204].

- **Smartphone Keys and Joysticks (legacy):** Presents approach where smartphone's hardware keys or joysticks enabled spatial interaction. We label this as a legacy approach, given the rarity of current smartphones featuring hardware keys or joysticks. Nonetheless, it offers valuable insights for a historical review of such techniques.

2.2.4 Motivation Behind Tracking Systems

The compelling concept of "bring-your-own-device" (BYOD) [19], which facilitates immediate interactions, ensures that the focus on inside-out smartphone tracking continues to be a significant area of interest for many researchers exploring distant displays [234]. Researchers are pushing the boundaries of embedded smartphone technology to develop innovative 2D [172] and 3D [80, 201] interaction techniques, thoughtfully navigating between the limitations of existing tracking hardware only to eliminate the need for instrumenting the environment with cameras and buying or wearing dedicated devices. Within the hybrid tracking category, researchers have not primarily focused on the accessibility of tracking technology, often employing hardware setups that are not easily accessible to the mainstream audience. Nonetheless, these studies offer valuable insights into the performance and user experience of various uni-modal and novel multimodal interaction techniques for distant displays, providing important findings [180, 250].

2.2.5 Tracking Various Input Modalities

Using a smartphone solely to facilitate interaction with a distant display enables users to employ their fingers for touch inputs and movement of their device-holding hand for device motion inputs. The input modalities enabled by an inside-out smartphone are:

- Touch input, enabled by the device's touchscreen for touch inputs [172]
- Device tilting input, enabled by device's orientation tracking [58, 200]
- Device translation input, enabled by device's translation tracking [18, 35]
- Device pose input, as simultaneous translation and orientation tracking [12]

While touch input and device tilting have been predominant in both past and recent research, there is an emerging trend towards utilizing smartphone cameras with simultaneous localization and mapping algorithms (SLAM) [65], which are natively implemented in modern smartphones via platforms like Google's ARCore [76] and Apple's ARKit [6]. These works illustrate how SLAM can be employed to precisely track the absolute world-space pose of the smartphone, particularly the motion of the user's device-holding hand, in distant display scenarios [12]. Nonetheless, approaches that expand the currently known device-tracking inputs to include body-tracking, head pose, and eye-gaze of the user remain mainly unexplored.

Hybrid setups have been explored in numerous studies, utilizing smartphone pose tracking [28, 142] by room-based cameras [292] or smartphone-attached markers or tracking devices [177, 226]. Additionally, they have facilitated many body-tracking-based input modalities, such as head pose [180, 250], body pose [113, 169] and eye-gaze [127, 259], by using glasses- [259] or room-based trackers [151]. With calibration, these input modalities were also tracked in absolute world-space to the distant display. In such hybrid settings, researchers often strategically distribute input data between various devices, optimizing for quality; for instance, while smartphone translation tracking via an outside-in system (e.g. depth camera) might be highly accurate, researchers note, that the smartphone's gyroscope might offer more accurate orientation tracking accuracy [59, 244]. While smartphones SLAM can adequately track the world-pose of the smartphone, making external tracking devices obsolete in that regard, outside-in systems continue to excel in enabling tracking of multiple user body parts [39]. Consequently, we believe that the exploration of hybrid setups, which leverage both smartphone pose tracking and various body-tracking input modalities through a combination of internal and external devices, underscores the potential and challenges in achieving precise and multimodal user interaction with distant displays. While the simultaneous tracking of all these modalities with a single smartphone presents a challenge, it signals an opportunity for additional research and exploration into the capabilities of smartphones.

2.2.6 Interaction Techniques

Selecting appropriate interaction techniques for inside-out systems poses a significant challenge, even to experts in the domain. The substantial portion of inside-out approaches investigated and compared the following interaction techniques, for applications ranging from 2D pointing [234], collaboration [80], to 3D object manipulation [263]:

- Touch [60, 263]
- Ray-casting [200, 234]
- Plane-casting [120, 201]
- Translation-based virtual-hand [12]
- Peephole interfaces [17, 33]

Due to the lesser input modalities that can be enabled by inside-out systems, researchers frequently have to compromise the efficacy of their interaction techniques, even though they acknowledge that alternative techniques might be more efficient but would require external trackers. For instance, Siddhpuria et al., in their study, could not find a feasible method to determine the absolute orientation of the smartphone in relation to the distant display using inertial measurement unit (IMU) [234]. Consequently, they needed to use "fixed origin" ray-casting [200], as opposed to the more preferable real-world absolute method. Such compromises create a captivating pull for an increasing hardware complexity and put researchers into a difficult decision between rich interaction techniques versus setup simplicity.

Hybrid configurations have demonstrated a variety of distant display use cases, wherein understanding the user's body position or gaze direction can facilitate highly efficient interaction techniques, such as ego-centric [207], head-pointing [180], or gaze+touch interactions [250, 259]. Additionally, hybrid tracking allowed researchers to investigate more complex user interfaces, such as 3D data visualizations, which require multiple input modalities that simultaneously enable many degrees of control, for 3D object translation, rotation, scale, and also 3D camera viewpoint controls [59, 153]. Moreover, a historical review of hybrid systems underscores the significance of the smartphone's touchscreen. Given the Midas problem [250], which is inherent in input modalities with a singular input state (such as hand, head, gaze, or body motions), the touchscreen has been essential in reliably segmenting these into meaningful interaction techniques. This has made smartphones an indispensable device, even in numerous systems incorporating external tracking. Within our taxonomy, we have distinguished the input modalities between the two tracking categories to show the difference in the role of touch in *inside-out smartphone only* and *hybrid* configurations. Primarily, inside-out systems are considerably more reliant on touch compared to hybrid systems, which emphasize more on device motion in 3D space, body interactions, and head or gaze inputs, utilizing touch predominantly for clicking and clutching.

Wigdor et al. argue that *"There can be no denying the power of the basics of the interaction language: get them right, and your system could seem simple, even "natural" to use. Conversely, get them wrong, and your user experience has no chance of success."* [284]. Frequently, user interfaces demand inputs across numerous degrees-of-freedom, necessitating not only the tracking of various input modalities but also their simultaneous tracking. These multimodal approaches demonstrate how we can significantly reduce the consecutive input states required, thereby simplifying the learning of interaction techniques. This is observable in works utilizing touch inputs for 3D object manipulations, where users employ up to three-finger gestures to gain full control over the 3D object and scene.

2.2.7 Hardware Approaches: Benefits and Limitations

Within inside-out systems, researchers challenged the boundaries of each new smartphone-adapted technology. They enabled inputs by using hardware keys [87], joysticks [35], inertial measurement unit (IMU) [58], touchscreens [172], cameras utilizing optical flow [18] or SLAM algorithms [12] to interact with the distant display. Researchers exploring hybrid setups have invested heavily in intricate in-lab hardware configurations to investigate spatial interactions with distant displays. For example, in a recent work, Hartmann et al. used six Kinect cameras (each connected to a PC) and a ten-camera Vicon motion tracking system for real-time tracking of the pose of the smartphone and the user's head, while the user was additionally wearing a hat with IR markers attached [90]. These are significantly complex hardware setups for tracking the smartphone and head pose, however very common among researchers in the domain.

Smartphone IMU Enabled Tracking

Many studies have shown that using the phone's movements enabled by the IMU, combined with touch, can outperform the use of many other alternatives, such as touch-only devices, mouse devices, Wii remotes (6DOF wand), especially in 3D object translation and rotation tasks [118, 256]. However, this requires that users calibrate their smartphone’s IMU to determine its correct orientation each time it is used [60, 226]. Moreover, to maintain input accuracy, repeated calibration, whether user-initiated or semi-automated, might be required during interaction. These studies demonstrate that while IMUs are generally sufficient for discrete motion detection, but are inadequate for precise and continuous position tracking necessary for 6DOF position and orientation tracking. Given that IMUs continually integrate acceleration with respect to time to calculate velocity and position (dead reckoning), any measurement errors, however small, accumulate over time [79, 232], leading to "drift" — an ever-increasing difference between perceived and actual device location.

Smartphone Camera Enabled Tracking

Rekimoto et al. [210] proposed an approach for simultaneously determining a smartphone’s position and orientation using printed 2D matrix markers (square-shaped barcodes on paper) attached to objects in the environment. Utilizing the appropriate phone application, such as ARToolKit [10, 175], the phone’s camera could locate and identify external markers to estimate its relative position and orientation. Based on this technology, researchers introduced "Sweep" and "Point-and-Shoot" interaction techniques, aimed at relative cursor control and target selection on distant displays [215, 264]. Additionally, they demonstrated techniques for 3D object manipulation [96], 3D mesh editing in 6DOF [95], bi-manual interaction [83, 275], and map interactions [215]. They also combined phone movement with touch input [164, 176] to manipulate AR objects displayed on the phone’s screen. Later, researchers investigated 2D interactions utilizing optical-flow analysis [33, 35], making environment markers obsolete. Wang et al. [280] introduced a compelling concept that enabled outdoor-use 2D gestures for phone control. Currently, modern smartphones can employ advanced computer-vision methods, such as dense SLAM [219] or similar technologies [68, 85], to identify the exact surface geometry of an unfamiliar environment and subsequently use it to estimate its own position and orientation. These solutions empower us to progress beyond local 6DOF tracking.

Smartphones can also be a powerful alternative to dedicated controllers. Utilizing the IMU, phone-orientation (3DOF) has been used to point at distant screens [38, 60] and in VR [61, 149] by using a ray-casting interaction technique. Moreover, the IMU has been utilized for selecting and manipulating objects in 3D environments using 2D rotational planes [80, 263] , enabling throw or swing gestures to facilitate media transfer between devices [58, 189], and recognizing uni-stroke letters [1] or symbols [123, 288] in combination with acceleration data.

2.3 Summary

We can see how the approaches underlying the classification of tracking technology into *inside-out* and *hybrid* setups significantly impact smartphone-based distant-display interaction across various universal design attributes and usability aspects. No existing works were able to bridge the gap between the two research streams by offering an inside-out solution that satisfies both the number of tracked input modalities as well as input precision. Consequently, utilizing a smartphone to facilitate distant display interactions predominantly remains confined to device-tracking inputs, such as touch and device tilting, occasionally extending to smartphone transition-based inputs, but generally excluding body-tracking inputs.

Based on our analysis, we believe that touch inputs, along with device (hand), head, gaze, and body pose tracking, should constitute the fundamental suite of input modalities for spatial interaction with distant displays. The capability to track all these input modalities, solely with an off-the-shelf smartphone and without additional external devices, would enable many powerful spatial interactions, discovered over decades of in-lab research (e.g. head or gaze pointing, virtual-hand or peephole interactions, body-centered inputs) for many other users, ideally for everyone. Similarly, to how the integration of accelerometers and touchscreens into smartphones expanded the domain of repurposed smartphones in many scenarios (from distant display to mixed reality), the adoption of new world- and face-tracking algorithms by modern smartphone has the potential to further notably amplify these advancements. The first steps were already made by using smartphone-based world-tracking in the domain of handheld AR [272, 273] , head-mounted displays [174] as well as using face-tracking in the domain of cross-device [267] and on-phone interactions [266]. The use of simultaneous world- and face-tracking on off-the-shelf smartphones has yet to be fully explored, as the initial instances of this technology have only recently been showcased in handheld AR scenarios [168]. In terms of interaction techniques, we can note that over the past, once the hardware constraints of a previous technology were replaced with a newer one, the fundamental set of spatial interaction techniques remained consistent: a handheld device for precise touch and motion tracking of the hand, head and body movements, independently of which domain was addressed (i.e. head-mounted VR/AR [50, 174], 3D interfaces on distant display [250], cross-device interaction [207]).

Pursuing Usability and Universal Design-Oriented Approaches in Spatial Interaction for Distant Displays

Given our "design for everyone" thesis philosophy, which emphasizes *universal design*, *usability*, and *application flexibility*, and resulting from our analysis of related work, we can intertwine the design philosophy with key limitations specific to the domain of spatial interaction with distant displays, as revealed by various related works. Consequently, we introduce two essential design directions that, when addressed, possess the potential to significantly improve spatial interaction with distant displays from a universal design standpoint:

- **Interaction Techniques: Maximize the number of input modalities**

  When enabled, interaction designers can freely choose and construct spatial interaction techniques, while fully prioritizing the technique's usability for the specific context. This allows even designers to transfer well-established and standardized interactions from related contexts or opt for techniques whose usability has been validated through research.

  When not enabled, interaction designers are constrained to use only a limited set of input modalities to construct necessary spatial interaction techniques. This prevents the implementation of the most suitable or standardized interaction technique for particular contexts. Frequently, it also imposes an additional cognitive burden on users, as the newly created interaction techniques necessitate user learning.

- **Hardware Solution: Minimize the need for outside-in tracking devices**

  When enabled, many users can effortlessly access new technology or utilize spatial interaction with a distant display, eliminating the need to purchase, set up, or maintain dedicated devices over time.

  When not enabled, spatial interaction with distant displays becomes accessible only to a specific user group, making it inaccessible to many other potential users, disciplines, and usage contexts.

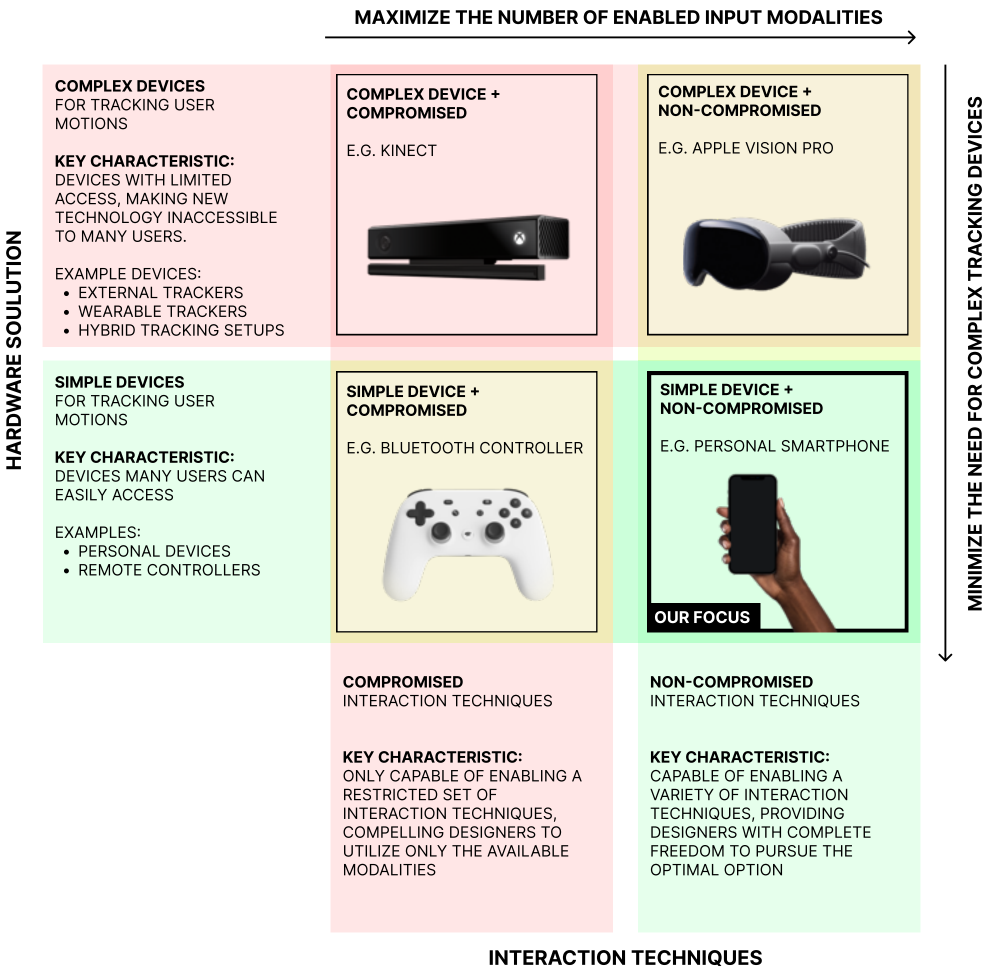


Figure 2.12: Four emergent design directions arise from the interweaving of the universal design philosophy and key limitations identified in related work on spatial interaction devices for distant display interaction. The resulting quadrants show the benefits and constraints among various approaches that may be considered when approaching spatial interaction for distant displays, grounded in the foundational hardware solution and the consequent interaction techniques and capabilities.

Four emergent design directions arise from the interweaving of the universal design philosophy and key limitations identified in related work on spatial interaction devices for distant display interaction. The resulting quadrants show the benefits and constraints among various approaches that may be considered when approaching spatial interaction for distant displays, grounded in the foundational hardware solution and the consequent interaction techniques and capabilities.

These two design directions converge into four quadrants, as depicted in Figure 2.12. From these quadrants, we can identify the trajectory of our work and extract concrete steps to follow our usability and universal design-oriented approach for spatial interaction with distant displays, concerning:

- **Interaction techniques**, we enable simultaneous, world-scale tracking of the essential input modalities required to assemble rich spatial interaction techniques. This includes the user's touch, hand, head, body, and eye-gaze inputs.
- **Hardware solution**, we propose an inside-out smartphone-based alternative with sufficient and comparable input capabilities.

In the following chapters, we present several ideas, projects, and prototypes that adhere to these design directions.

Parts of the following Chapter 3 have been published as:

Teo Babic, Harald Reiterer, and Michael Haller. 2018. Pocket6: A 6DoF Controller Based On A Simple Smartphone Application. In: Proceedings of the Symposium on Spatial User Interaction - SUI '18. New York, New York, USA: ACM Press, 2018, pp. 2–10. DOI: https://doi.org/10.1145/3267782.3267785

The responsibilities of this publication were as follows: I formulated the research questions, designed and implemented the research prototype, designed and conducted the user study, analyzed the study data, and wrote the paper. Harald Reiterer and Michael Haller supervised the work.

Supplemental Material

The QR codes below allow you to access the supplemental material by either scanning it using a mobile phone (print) or by clicking on it (digital).

Paper

Conference Talk

Repository

3

Repurposing the Smartphone into a Spatial Tracker and Controller

Handheld input controllers with external 6DOF tracking have become the benchmark for 3D interaction. Users typically find them very effortless to use, since their translation and rotation-based movement tracking allows them to be used as "virtual hands" to locate, grab, and manipulate virtual 3D objects — similarly as in the real world. However, when external tracking is not possible (often due to infrastructure or mobility constraints), interaction possibilities become drastically reduced and highly dependent on other available devices and their features. Previous studies have investigated the potential of using the most convenient and readily available device for end-users, their personal smartphone, to facilitate 2D and 3D interactions. However, these approaches typically required additional external hardware alongside the smartphone (such as cameras, trackers, or printed markers) to make the smartphone spatially-aware. The initial challenge of enabling spatial awareness in smartphones led researchers to suggest interaction methods limited to 3DOF, as it was simpler to implement than the full 6DOF input [201, 234]. Therefore, our aim was to investigate the possibility of developing a smartphone application capable of overcoming these limitations.

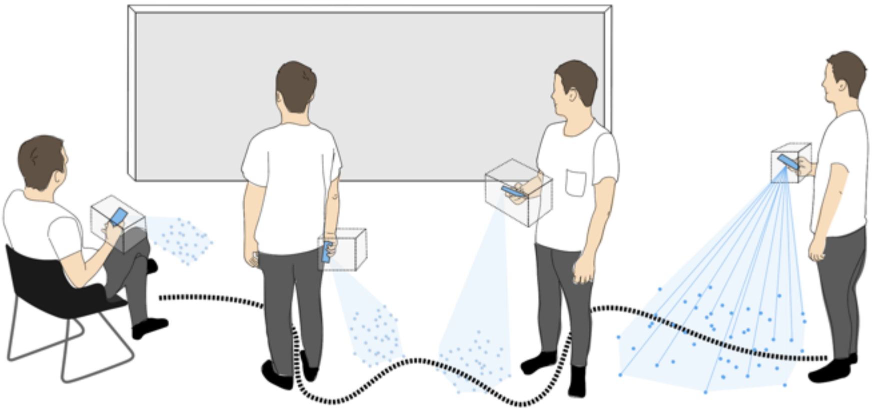


Figure 3.1: An illustrative overview of the Pocket6 smartphone application. The figure showcases the application's utilization of modern smartphones' spatial tracking capabilities for user interaction with distant displays. It highlights the simplicity of the system, which requires no external trackers, markers, or wearables. The figure also visually represents user interactions in various postures, such as standing, sitting, and hands-down, from a subtle control space that automatically re-calibrates based on the user's orientation, position, and posture changes.

In this chapter, we present, implement, and evaluate Pocket6, a smartphone application utilizing modern smartphones' AR tracking capabilities to enable 6DOF user input without external hardware. Our application stands out for its simplicity and accessibility, needing just a standard smartphone and no external trackers, markers, or wearables. Our novel inside-out tracking approach allows users to comfortably perform subtle 3D inputs everywhere in world-scale, without any spatial or postural limitations. Due to our algorithm for auto-calibration of the control space, users can freely adjust their position, orientation, or posture during interacting, as depicted in Figure 3.1. Given the reduced spatial and postural constraints, users can interact while standing, sitting, or while having their hands down by their sides. We evaluate the performance of our approach and compare it against a high-end VR controller. We demonstrate that Pocket6 enables users to perform precise subtle gestures within a control space of only 16 × 9 × 16 cm. Additionally, we evaluate how body postures *standing*, *sitting*, and *hand-down*, impact user performance with the application. Finally, we showcase the use of Pocket6 in a wide range of applications for 2D and 3D object manipulation, underscoring its suitability for diverse real-world scenarios.

3.1 The Pocket6 Concept

Pocket6's foundational concept draws from Apple's Augmented Reality toolkit, ARKit [6]. ARKit utilizes an inside-out mobile tracking method, leveraging a technique termed SLAM (simultaneous localization and mapping [65]) sensor fusion. It uses data from the smartphone's existing camera and IMU sensors to calculate where the device is in world-space and translates the physical location and movement into a movement within a virtual space or environment. Typically, these AR toolkits are being used to facilitate handheld AR experiences on smartphones, anchoring and tracking virtual objects in the real world. A few examples can be seen in Figure 3.2.

Figure 3.2: Two examples of augmented reality enabled by smartphone's inside-out tracking. On the left, an IKEA app allows users to position virtual furniture within their actual living space. On the right, the *Live Home 3D* app lets users design architectural and interior concepts for their homes.

Rather than fixing virtual objects to real-world coordinates, in Pocket6[1], we employ ARKit to anchor a virtual 3D control space, within which we can track the phone's position and orientation as shown in Figure 3.3, B. This configuration allows 2D and 3D inputs for external applications that can be run on a distant display, as depicted in Figure 3.3, G. We developed Pocket6 on Apple's iPhone ecosystem, it's compatible with ARKit-supported devices from iPhone 6s, till the latest iPhone 15 [6]. Comparable toolkits, like Google's ARCore [76], could also be used to enable this implementation. As Pocket6 served solely as an input application, we developed a Windows-based program to showcase output applications on a distant display. The application communicated with Pocket6 via the local wireless network, sending the smartphone's motion updates at 60 Hz.

---

[1] Github - Pocket6 implementation: https://github.com/teobabic/pocket6

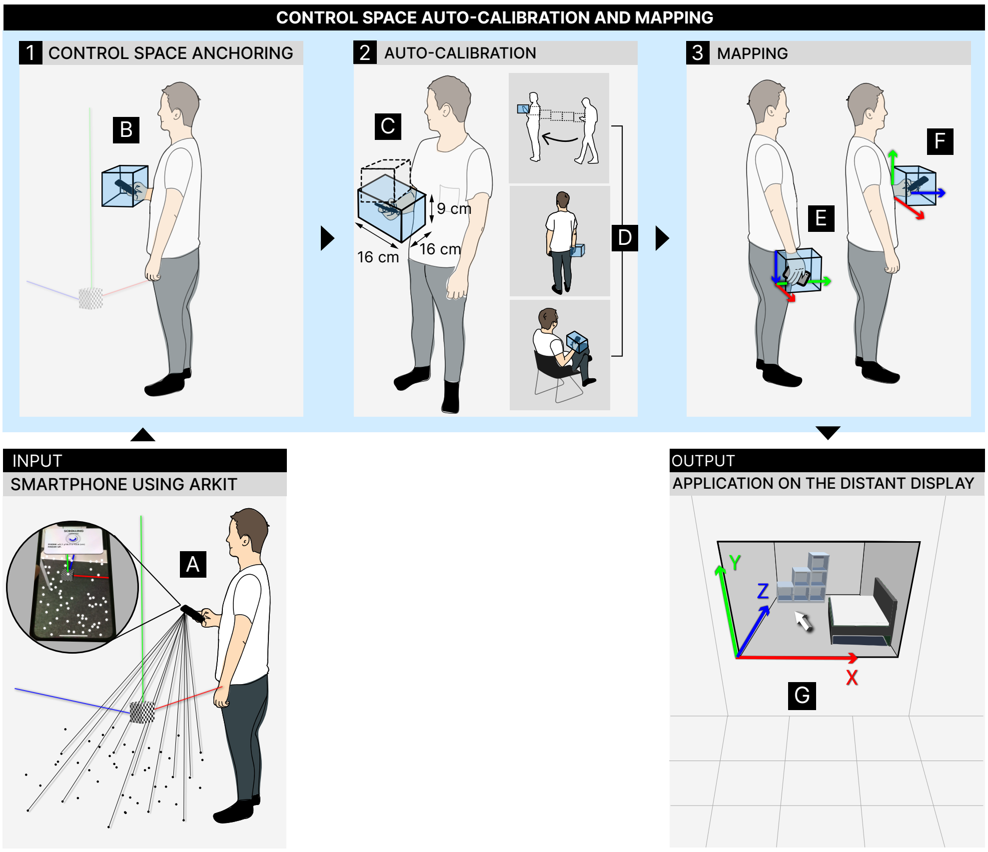


Figure 3.3: Proposed smartphone app-based approach for subtle, 6DOF, world-scale inputs. A user interacts within a virtual 3D control space (visualized as a bounding box). The coordinates of the real interaction space are mapped automatically to the digital screen. The movement of the cursor corresponds with the smartphone's motions.

3.1.1 Three-step Control Space Calibration and Mapping Algorithm

One of the primary objectives of our implementation was to eliminate the necessity for users to initiate calibration procedures. To achieve this, we implemented an auto-calibration algorithm that re-calibrates the control space position and rotation whenever users:

- Open the application for the first time, cf. Figure 3.3, A and B,
- Move their hand to a different position, or

- Change their posture or walk around, cf. Figure 3.3, C and D

Hereby, in every scenario, the control space consistently *follows* the user's motion, consequently enclosing the user's hand at all times. In detail, this algorithm involves three distinct steps:

1. **Control Space Anchoring:** Firstly, using the 3D Cartesian coordinate system in real-world metric units (cm) set by ARKit cf. Figure 3.3 A, we align the control space's geometric center with the smartphone's position cf. Figure 3.3 B, and align their (z—axis) rotations cf. Figure 3.3 C. The boundaries of the control space were determined by the results of an empirical study performed during its development, which suggested that the control space should reflect the aspect ratio where 1 cm of the control space represents 120 px of the distant display. In our setup, a distant display screen resolution of 1920 × 1080 px is mapped to a control space measuring 16 × 9 cm (along the x— and y—axis). For 3D applications, we set a depth to the same size as the application's width provided by metric units of the game engine; this translated to a 16 cm depth in the control space (on the z—axis), as shown in Figure 3.3 C.
2. **Re-Anchoring:** If the phone's position moves outside the control space boundaries, as shown in Figure 3.3 C and code-snippet below[2], we assume that the user has changed their position, initiating the re-anchoring. Giver our objective on user's motion freedom, our approach also allows users to interact with the controlling hand pointing in a relaxed downward position [152], the axes of our control space must adapt accordingly, as illustrated in Figure 3.3 E. This occurs automatically once the phone detects an upside-down orientation (i.e. the phone's x—axis rotation is between 130° and 230°).
3. **Mapping:** Finally, we normalize the smartphone's position within the control space boundaries and forward its data as position, rotation, and touchscreen events to an application, for example, to move a 3D cursor, as depicted in Figure 3.3 G.

---

[2] Github - Pocket6 algorithm implementation: https://github.com/teobabic/Pocket6/blob/master/Pocket6%20Smartphone%20App/Assets/Scripts/Pocket6Logic.cs

Listing 3.1: Code snippet of the control space remapping procedure

```
float allowedOffset;    // Allowed offset before remapping (in cm)
Vector3 deltaDistance;  // Distance between phone and control space center

void Update()
{

 ...

 if (
 Mathf.Abs(deltaDistance.x) > (ControlSpaceSize.x / 2) + allowedOffset ||
 Mathf.Abs(deltaDistance.y) > (ControlSpaceSize.y / 2) + allowedOffset ||
 Mathf.Abs(deltaDistance.z) > (ControlSpaceSize.z / 2) + allowedOffset
 )
 {
    RemapControlSpace();    // Remap
    return;
 }

 // Do cursor mapping and other interaction functions
 ...

}

void RemapControlSpace{}
{
    controlSpace.position = smartphone.position;
    controlSpace.eulerAngles = new Vector3(0, smartphone.eulerAngles.y, 0);
}

...
```

To minimize both jitter and latency, we used the *1 EURO* filter [44, 45].

3.2 Evaluation

We conducted a three-part empirical study to explore the benefits and limitations of our proposed app-based tracking approach. We compared Pocket6 to a high-end, state-of-the-art, VR controller. Furthermore, we investigated how well participants can perform input from different body postures. All three experiments were conducted with the same techniques, participants, and apparatus.

3.2.1 Conditions

We tested four distinct conditions (refer to Figure 3.4) to compare the performance, accuracy as well as subjective user feedback of our proposed condition:

- **In-Front**: In this setup, participants stood and held the smartphone naturally *in-front* of their torso (cf. 3.4 B).
- **Sitting**: While *sitting* on a chair, participants rested their elbow on the armrest and held the smartphone above their waist (cf. 3.4 C).
- **Hands-Down**: participants stood and held the smartphone in a *hands-down* posture (cf. 3.4 D).
- **Baseline**: for the baseline, participants stood and used an externally-tracked VR controller, holding it in front of their torso (cf. 3.4 A).

3.2.2 Apparatus

The study was conducted in a quiet room, where the participants stood/sat two meters away from the display, a 1920 × 1080 pixels (32”) Samsung TV, which showed all the tasks. For the *Baseline* condition, we used an HP Microsoft Mixed Reality headset along with its accompanying Windows Mixed Reality motion controller, as shown in Figure 3.4 A. As the headset itself has built-in controller tracking cameras, we mounted it on a tripod in front of the participant to effectively capture their hand movements, approximately at a distance 0.5 to 0.7 m away. We adjusted the headset’s height to align with each participant’s chest level. The raw data received from the headset controller and Pocket6 were processed through exactly the same signal filtering, control space remapping, and mapping algorithms.

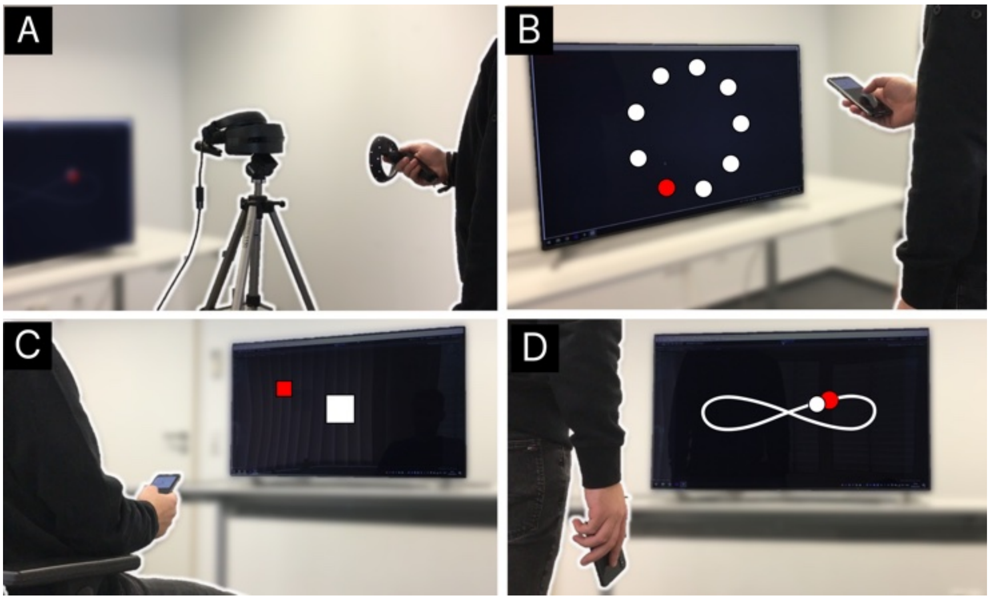


Figure 3.4: Illustration of the experimental conditions and tasks. Conditions: A *Baseline* participants stood and used an externally-tracked VR controller, holding it in front of their torso. B *In-Front*, participants stood and held the smartphone naturally in front of their torso. C *Sitting*, participants sat on a chair, resting their elbow on the armrest and holding the smartphone above their waist. D *Hands-Down*, participants stood with the smartphone in a hands-down posture. Tasks: B 2D Fitts' law task, C 3D placement task, and D 2D tracking task.

3.2.3 Participants

A group of 12 paid volunteers (5 females, 7 males) from the local company, aged between 20 and 42 years ($M = 31.5$, $SD = 5.17$), took part. Every participant was right-handed, had no prior experience with 6DOF input controllers, and used smartphones on a daily basis.

3.2.4 Study Design

We employed a repeated measures within-subject design. Participants' performance was investigated across three tasks, as depicted in Figure 3.4:

1. 2D tracking task,

2. 2D Fitts' law task, and
3. 3D placement task.

For each task, we carried out an experiment, with detailed explanations provided in the subsequent dedicated chapters for each experiment. These three tasks were built on each other with an increasing input complexity and were performed in sequence. As a result, participants gradually practiced as tasks transitioned from simpler to more complex ones. In the beginning, each participant was welcomed and introduced to the experiment procedure. After they completed a background questionnaire, participants were given 5 minutes time to familiarize themselves with the conditions. We recorded performance data via computer logs and collected subjective feedback through an exit questionnaire.

3.2.5 Study 1: Path Tracking in 2D

The purpose of the first experiment was to use a tracking task to determine if different input conditions affected users' precision in guiding their hand in mid-air. In fact, we wanted to find out how accurately participants could continuously follow a moving target with the cursor, even when the target did fast directional changes.

Our task was based on the ideas of [46, 240], where participants had to trace a moving target on a pre-defined path, as shown in Figure 3.5. Following [46], we selected a narrow 720 × 270 px eight-shaped path (∞) with a target movement speed of 2π seconds, ensuring a continuous hand motion setting the duration of one trial (lap). We chose an eight-shaped path since it is more complex compared to a circular or elliptical path and required a more fine-granular input from the participants. Each trial began and ended when the moving target crossed the bottom part of the left loop. From there the target was moving clockwise on the left loop and counterclockwise on the right. Both the cursor controlled by participants as well as the target had a 50 px radius.

Every participant completed seven trials under each of the four conditions. Participants began with two practice trials and then proceeded to five evaluated trials. In total, each participant evaluated 20 trials. During the experiment, we logged the Cartesian distance in pixels (px) between the path and cursor at a rate of 60 Hz.

Results

Error rate All traces for each condition from all participants are illustrated in Figure 3.5. The average errors recorded were as follows: *Baseline* condition at 68.60 px ($SD = 15.65$), *In-Front* at 73.10 px ($SD = 20.40$), *Sitting* at 73.97 px ($SD = 21.26$), and finally *Hands-Down* at 90.10 px ($SD = 29.91$). Using a one-way repeated measures ANOVA with α set at .05, we found significant differences among the four conditions ($F_{3,33} = 8.152, p < 0.001$). As the collected data did not violate the assumption of sphericity, no corrections were necessary. Using Bonferroni corrections for post-hoc pairwise comparisons, we observed that the *Hands-Down* condition had a notably higher inaccuracy (averaging between 16 to 21 px) than all other conditions (*Baseline* $p = 0.001$, *In-Front* $p = 0.014$, *Sitting* $p = 0.002$). The differences between the *Baseline*, *In-Front*, and *Sitting* conditions were not significant, averaging a discrepancy of less than 6 px.

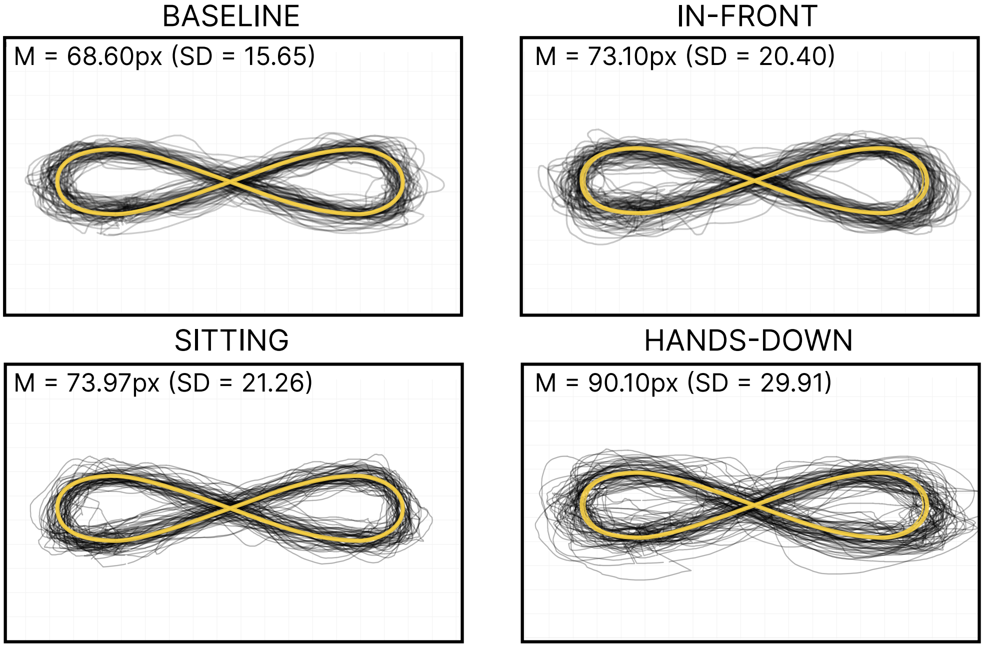


Figure 3.5: Visualization of cursor traces from all participants' tracking tasks across the four conditions. The target's eight-shaped trajectory is highlighted in yellow.

The Pocket6 conditions, *In-Front* and *Sitting*, enabled participants to perform continuous hand movements with the same granularity and precision as the externally tracked *Baseline* controller. Furthermore, we can see that in the *Hands-Down* condition, participants performed less accurately than in all other conditions.

3.2.6 Study 2: Pointing in 2D

The goal of the second experiment was to evaluate the performance of 2D pointing and clicking. This experiment was based on the ideas of [152, 268], we used a 2D Fitts' law task [106, 218], which is accessible on [157]. In this experiment, participants had to point and click a set of circular targets displayed on the distant display. A trial was considered successful when the initial *click-down* and subsequent *click-up* actions took place within the target's boundaries. Participants needed to successfully select each target to advance to the next trial. We evaluated two amplitudes (400 and 800 px) and three target widths (50, 100, and 200 px), resulting in an Index of Difficulty (ID) spanning from 1.6 to 4.1 bits. In each of the four conditions, participants completed a practice block with 3 targets, succeeded by a randomized block of 9 measured trials. Every block contained all possible combinations of amplitudes and target widths. In total, each participant produced 216 data points, derived from 4 conditions × 2 amplitudes × 3 widths × 9 target selections.

Results

We analyzed the main effects of our conditions, focusing on standard metrics like throughput, error rate, and movement time. For our analysis, we employed a repeated measures ANOVA ($\alpha = .05$) and conducted pairwise tests, applying Bonferroni corrections for post-hoc analysis.

Throughput For the conditions, *Baseline* recorded a throughput of 1.86 bps ($SD = 0.59$), *In-Front* achieved 1.82 bps ($SD = 0.55$), *Sitting* 1.83 bps ($SD = 0.52$), and *Hands-Down* noted 1.36 bps ($SD = 0.50$). A one-way ANOVA showed a significant difference among the four conditions ($F_{3,15} = 8.00, p < 0.006$). Due to the violation of the sphericity assumption, we report the Greenhouse-Geisser adjusted values. The post-hoc test showed that *Hands-Down* had a significantly lower throughput by an average of 26% when compared to all other conditions: *Baseline* ($p = 0.027$), *In-Front* ($p = 0.01$), and *Sitting* ($p < 0.001$). Other pairs did not differ significantly; their throughput differed on average for less than 2%.

Error rate The average error rate for *Sitting* was 8.3% ($SD = 10.59\%$), *In-Front* 11.26% ($SD = 13.54\%$), *Baseline* 12.19% ($SD = 12.98\%$), and *Hands-Down* 18.05% ($SD = 15.31\%$). A one-way ANOVA showed a significant difference between the conditions ($F_{3,33} = 9.720, p < 0.001$). There was no violation of the sphericity assumption. The post-hoc test showed that *Hands-Down* had a significantly higher error rate compared to all other conditions: *Baseline* ($p = 0.035$), *In-Front* ($p = 0.018$), and *Sitting* ($p = 0.001$). Besides the *Hands-Down*, there were no significant differences in error rate among the other conditions. An overview of the results can be found in Figure 3.6.

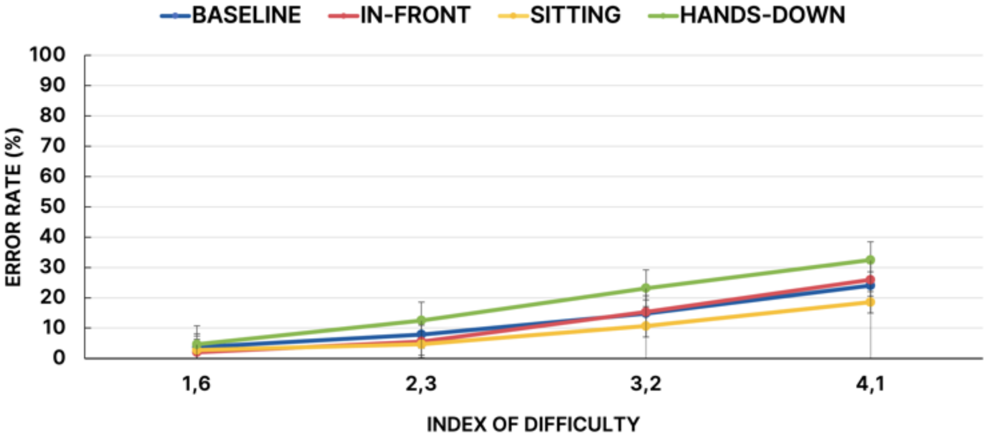


Figure 3.6: Error rate distribution for each *input device* condition, broken down by individual Index of Difficulty (ID) values.

Movement time In terms of movement time (MT), the *Baseline* condition had an average MT of 1508 ms ($SD = 447$), *In-Front* 1735ms ($SD = 658$), *Sitting* with 1712 ms ($SD = 602$), and *Hands-Down* with 2148 ms ($SD = 933$). A one-way ANOVA revealed a significant difference in MT across the conditions ($F_{3,33} = 17.5931, p < 0.001$). The post-hoc test found that *Hands-Down* had a significantly higher MT compared to all other conditions: *Baseline* ($p = 0.002$), *In-Front* ($p = 0.001$) and *Sitting* ($p = 0.002$). No significant differences were observed among the other condition pairs. Figure 3.7 provides an overview of these results.

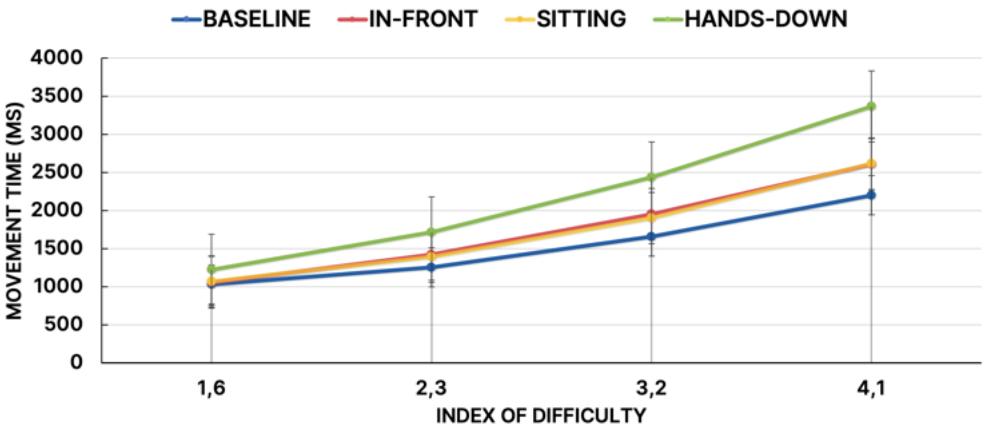


Figure 3.7: Movement time distribution for each *input device* condition, broken down by individual Index of Difficulty (ID) values.

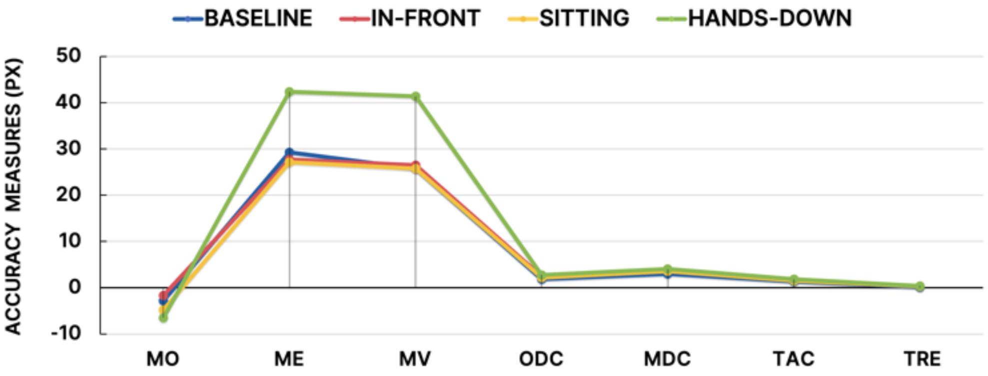


Figure 3.8: Accuracy metrics of the Fitts' law task across each *input device* condition. The measures include Movement Offset (MO), Error (ME), Variability (MV), Orthogonal Direction Change (ODC), Movement Direction Change (MDC), Task Axis Crossing (TAC), and Target Re-Entry (TRE).

Fitts' law accuracy measures The accuracy measures of the Fitts' law task are depicted in Figure 3.8 [156]. The results indicate a high similarity among the *Baseline*, *In-Front*, and *Sitting* conditions. We also found that the primary limitations of the *Hands-Down* condition were attributed to movement error (ME) and movement variability (MV). This indicates that in the *Hands-Down* condition, participants' cursor movement towards the target was much further away from the ideal straight path (ME) and that their movements were also less consistent between the trials (MV) compared to the other conditions. Finally, we can also report a weak correlation ($R^2 < 0.3$) between movement times and IDs across all conditions.

We can see that our Pocket6 conditions, *In-Front* and *Sitting*, enabled users to achieve comparable 2D pointing and clicking performance in both speed and accuracy to the externally tracked controller, *Baseline*. However, the *Hands-Down* condition lagged behind, exhibiting slower speeds and inaccuracy compared to all other conditions.

3.2.7 Study 3: Object Manipulation in 3D

In our third experiment, we evaluated 3D interaction using a 3D placement task, focusing solely on translation, based on the experiment of Vuibert et al. [271]. The objective was to determine whether all four conditions provide the same accuracy and speed when participants perform continuous stop-and-move 3D hand movements in mid-air.

Participants were presented with two squares on the screen. They were tasked with aligning these squares using a 3D drag-and-drop hand motion. In each trial, participants maneuvered their cursor along the x— and y—axis to grab a white movable square using a tap-and-hold gesture, as depicted in Figure 3.9 A, B and C. After grabbing the square, their objective was to match its position and dimensions with a red reference square, cf. Figure 3.9 D. Participants could re-size the dragged square by moving their smartphone on the z—axis of the control space, cf. Figure 3.9 E. Moving the smartphone towards the positive direction on the z—axis reduced the size of the draggable square, and the other way around. For x— and y—axis cursor movement, we used the default control rate of 120 px on-screen is 1 cm in control space; for square resizing, we opted for a finer mapping for square resizing, where a 15 px increase or decrease of the square's width corresponded to a 1 cm movement on the z-axis of the control space. When the two squares were correctly aligned, they both turned green, and the user needed to confirm by lifting the finger from the touchscreen, cf. Figure 3.9 F. For correction checking, we used a tolerance of 10 px for both the position and scale. In the case of a negative match, participants had to re-grab and re-do the alignment. Once the alignment was successful, both squares disappeared, and participants needed to move their 3D cursor back to the middle of the interaction area (x-, y- and z-axis), with the hint of a small cursor widget. Afterwards, the next trial was shown.

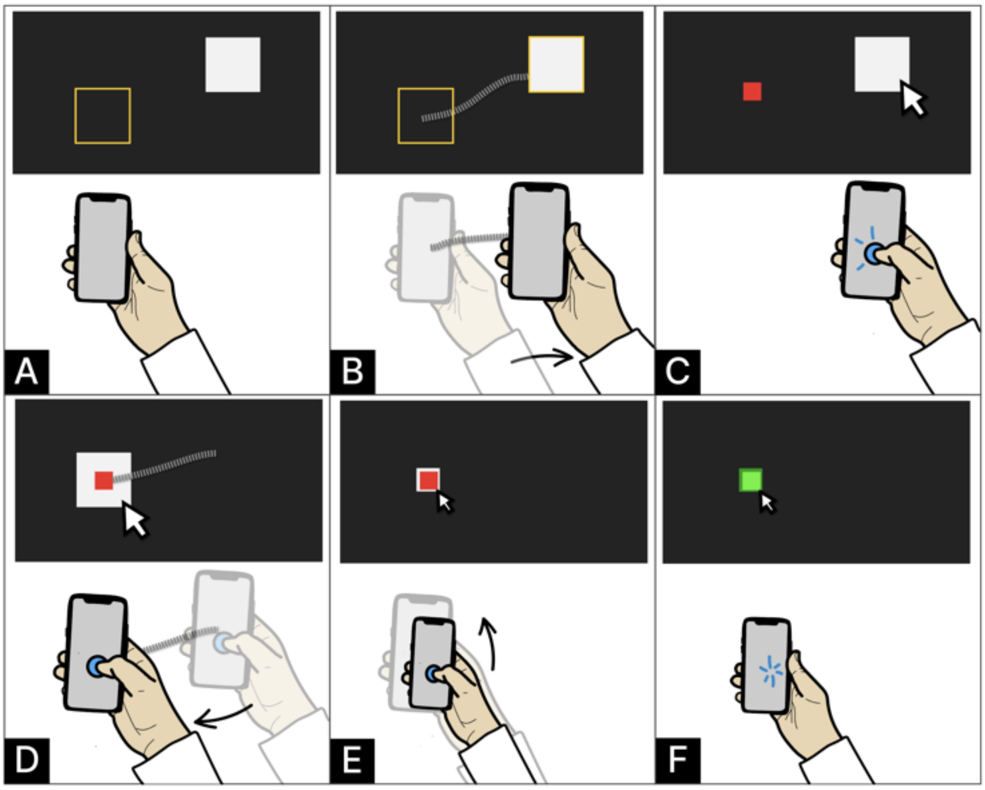


Figure 3.9: The sequence of user's actions in the 3D placing task. A-C: Participants use the yellow cursor to grab the white movable square with a tap-and-hold gesture. C: The goal is to align the grabbed square with the red reference square. D-E: Adjusting the size of the dragged square by moving the smartphone along the $x-$, $y-$, and $z-$axis of the control space. F: Successful alignment is indicated when both squares turn green, and the user lifts the finger from the touchscreen.

The target square's position, in terms of x and y coordinates, was determined by an amplitude (either 250 or 400 px) that described a radius from the screen's center, combined with a random angle that varied from 0 to $360°$. For defining the distance on the z-axis, we employed four distinct target dimensions (50, 95, 155, 200 px). At the beginning of each trial, the draggable square began with a size of 125 px, which also marked the starting position on the z-axis. Participants completed a practice block first, followed by a study block. Every block encompassed all possible combinations of amplitudes and target sizes. In total, each participant completed 64 trials, determined by 4 conditions × 2 amplitudes × 4 target sizes × 2 repetitions.

Results

Placement time We removed 2.1% outliers caused by technological errors and semantic errors, like participants trying to grab the target square instead of the draggable square. On average, the *Baseline* condition had a placement time of 2201 ms ($SD = 723$), the *In-Front* condition took 2334 ms ($SD = 913$), the *Sitting* condition 2212 ms ($SD = 861$), and the *Hands-Down* condition recorded 2543 ms ($SD = 935$), as shown in Figure 3.10. Using a one-way ANOVA with a significance level of α = .05, we found no significant difference in placement times across the conditions ($F_{3,33} = 5.620, p < 0.063$).

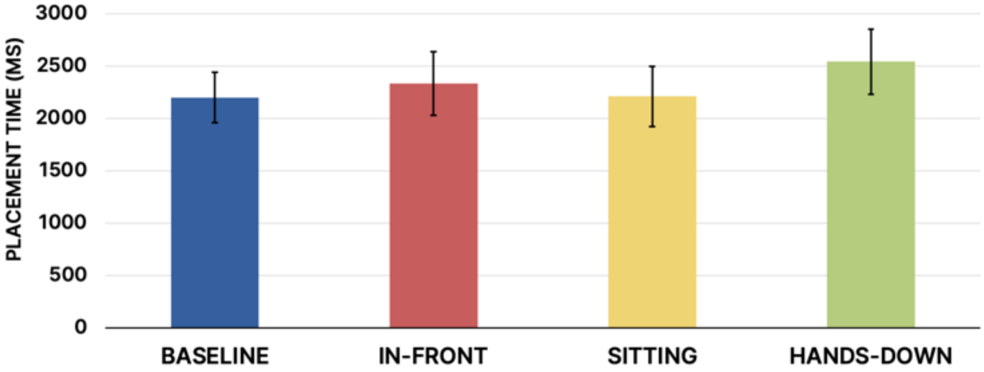


Figure 3.10: Placement time distribution for each *input device* condition.

Error rate With less than 1%, the error rate and its analysis were negligible since participants did not let go of the draggable square, which would count as an error until they got the indication that the alignment was correct.

In summary, we have shown that all Pocket6 conditions *In-Front*, *Sitting*, and *Hands-Down* performed comparably to the externally tracked *Baseline* condition.

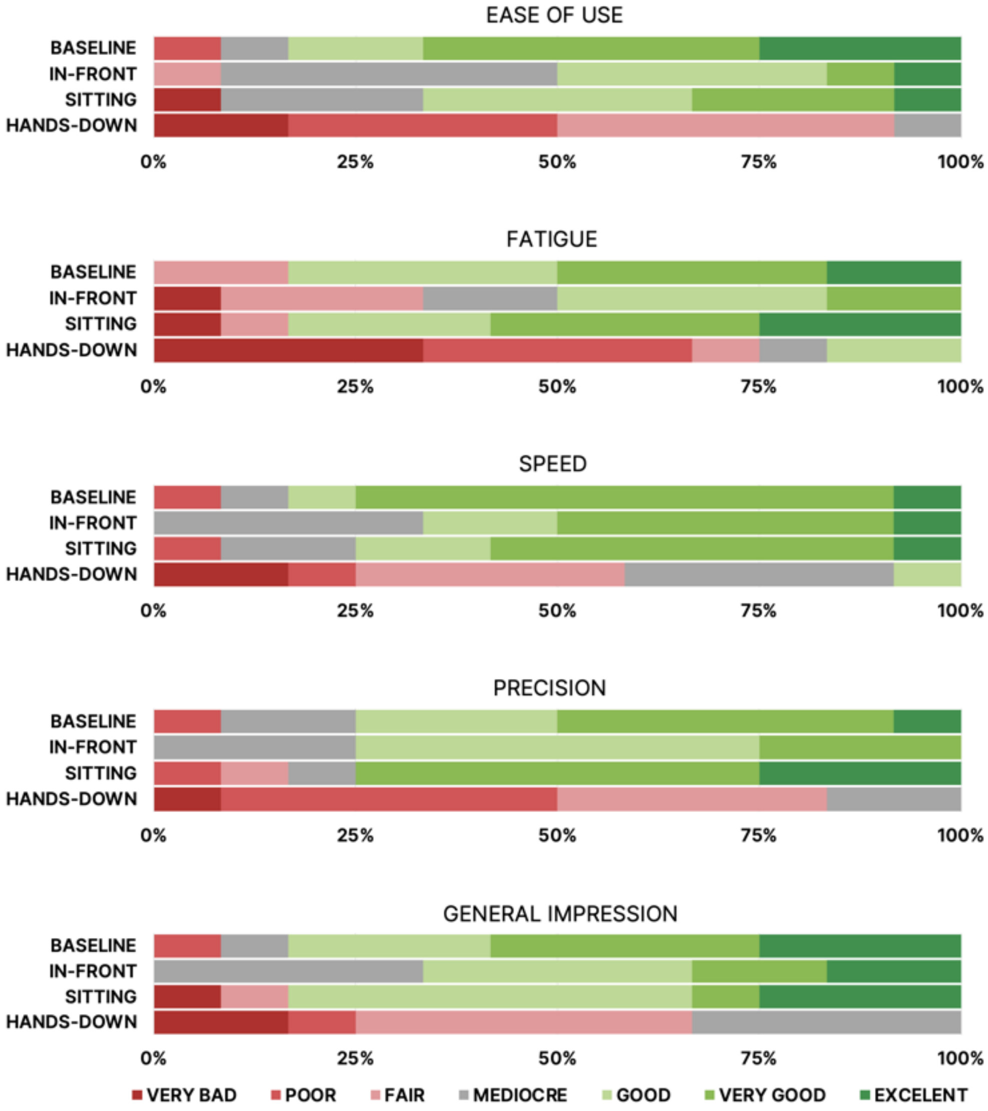


Figure 3.11: Participants' subjective feedback ratings for each *input device* condition, covering aspects such as ease-of-use, fatigue, speed, precision, and overall impression.

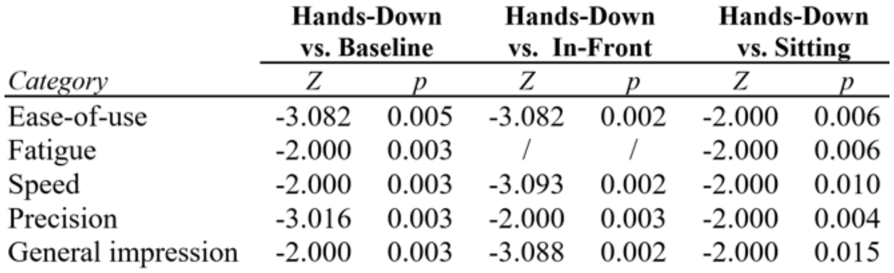
3.2.8 User Feedback

After participants had experienced all study tasks, we asked them to provide their unconstrained subjective feedback based on the overall experience across all three experiments, by rating each condition for *ease of use, fatigue, speed, precision*, and *overall impression* on a 7-point Likert scale, where a higher score indicated a better rating. The questionnaire can be seen in Appendix A.1. Finally, we asked participants for additional comments, suggestions, or recommendations. The goal was to learn from initial user reactions and comments with a special focus on exposing differences between our conditions.

Results

Questionnaire analysis Based on the Friedman test, there were significant variations in the ratings for ease of use ($\chi^2(3) = 20.00, p < 0.001$), fatigue ($\chi^2(3) = 18.00, p < 0.001$), speed ($\chi^2(3) = 19.00, p < 0.001$), precision ($\chi^2(3) = 19.00, p < 0.001$), and overall impression ($\chi^2(3) = 18.02, p < 0.001$) based on the condition being evaluated, as illustrated in Figure 3.11.

An in-depth analysis using the Wilcoxon Signed Rank test revealed no significant differences between the *Baseline*, *In-Front*, and *Sitting* conditions across all categories. However, a significant difference was observed between the *Hands-Down* condition and all other conditions across all categories, with the exception of *fatigue* when comparing *In-Front* and *Hands-Down*, as detailed in Table 3.1.



	Hands-Down vs. Baseline		Hands-Down vs. In-Front		Hands-Down vs. Sitting	
Category	*Z*	*p*	*Z*	*p*	*Z*	*p*
Ease-of-use	-3.082	0.005	-3.082	0.002	-2.000	0.006
Fatigue	-2.000	0.003	/	/	-2.000	0.006
Speed	-2.000	0.003	-3.093	0.002	-2.000	0.010
Precision	-3.016	0.003	-2.000	0.003	-2.000	0.004
General impression	-2.000	0.003	-3.088	0.002	-2.000	0.015

Table 3.1: Pairs with significant differences in the subjective feedback, as determined by the Wilcoxon Signed Rank test.

Participants general impression Finally, we delve into the feedback provided by participants regarding their experience with Pocket6 across different conditions. Feedback from participants indicated that Pocket6 surprisingly exceeded their expectations in performance. The majority of participants favored the *Baseline* and *In-Front* conditions due to their speed, accuracy, and user-friendliness. Participants reported that the ergonomics of a controller seems to be very important. They also pointed out that the iPhone X (174 g) felt somewhat heavier compared to the VR controller (171 g). Half of the participants agreed that the *Sitting* condition was very comfortable, especially because it allowed them to rest their elbows on the chair's armrest. However, they also noted that this is not always required and that it could also be a serious limitation. A few participants felt that resting the elbow made them lazy and that they did not want to raise their arm once rested, for example to reach for items at the upper side of the screen, where lifting the elbow might be required, this caused minor frustrations.

Specific comments regarding the hands-down posture From a subjective standpoint, all participants agreed that the *Hands-Down* condition was the most fatiguing and challenging to use. The primary reason for this was the need to precisely adjust the hanging down extended arm "lever" (the kinematic chain stretching from the neck to the fingertips). Moreover, many participants were unaccustomed to interacting with their hands positioned down beside their body. Participants believed that sustained interaction in a *Hands-Down* posture would be challenging for a long duration. However, there was a consensus that the axes chosen for cursor mapping in the *Hands-Down* condition were easy to get used to. A participant, *P3*, mentioned that the *Hands-Down* condition lacked hand-eye coordination as she couldn't see her hand while interacting.

Specific comments regarding the sitting posture Some participants noted that while *Sitting* and *In-Front* conditions performed equally good, each had its own minor drawbacks. With the *In-Front* condition, participants found it easier to overshoot targets, like in the Fitts' law tasks, but on the other side, it allowed for freer and unrestricted motions. Conversely, while the *Sitting* condition appeared more comfortable, due to the resting elbow, some participants perceived it as inconvenient. Some participants explicitly disliked the *Sitting* condition, explaining that it was too limited due to the elbow rest. Participants generally favored having complete freedom of motion to position the control space as they desired. They felt that the *In-Front* condition offered adequate comfort and precision without needing an armrest.

3.3 Design Recommendations

Participants noted that the ergonomics of larger smartphones, like the iPhone X, feel as if they were holding a heavier device. While this was perceived as a minor concern, we suggest some simple add-on enhancements to mitigate this issue. Enhancing the phone grip can be achieved by utilizing a readily available smartphone cover equipped with extra handles, as illustrated in Figure 3.12.

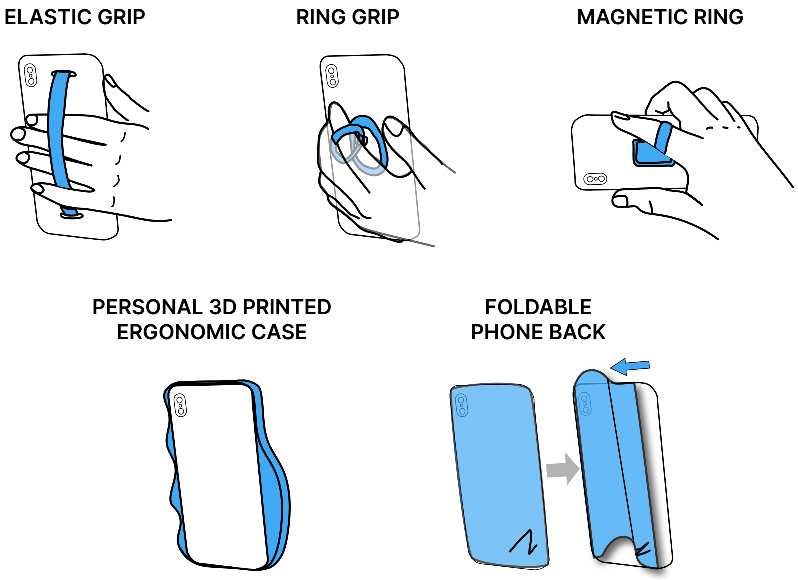


Figure 3.12: Proposed solutions for improving the ergonomics of a smartphone.

3.4 Applications

In this section we demonstrate how Pocket6 can immediately be used to control a wide variety of real-world applications, including Google Earth, YouTube, a text editor, PowerPoint, a furniture rearranging tool. The showcased applications aim to combine subtle mid-air gestures with simultaneous touch events to create interactions in convincing real-world scenarios. Pocket6 enables control of both 2D and 3D cursors, through a combination of subtle mid-air gestures, such as spatial translation and rotation, alongside touch inputs like taps and long-press events.

Maps

The first application area is maps, as illustrated in Figure 3.13. To begin, users manipulate the 2D cursor over the map by translating the smartphone through the air. Zooming in or out on the map is achieved by swiping a finger upwards or downwards on the touchscreen. Since these input actions (2D cursor movement and continuous zoom) are performed simultaneously, and because the map zooms into the location of the cursor, this allows for very efficient and precise user interaction.

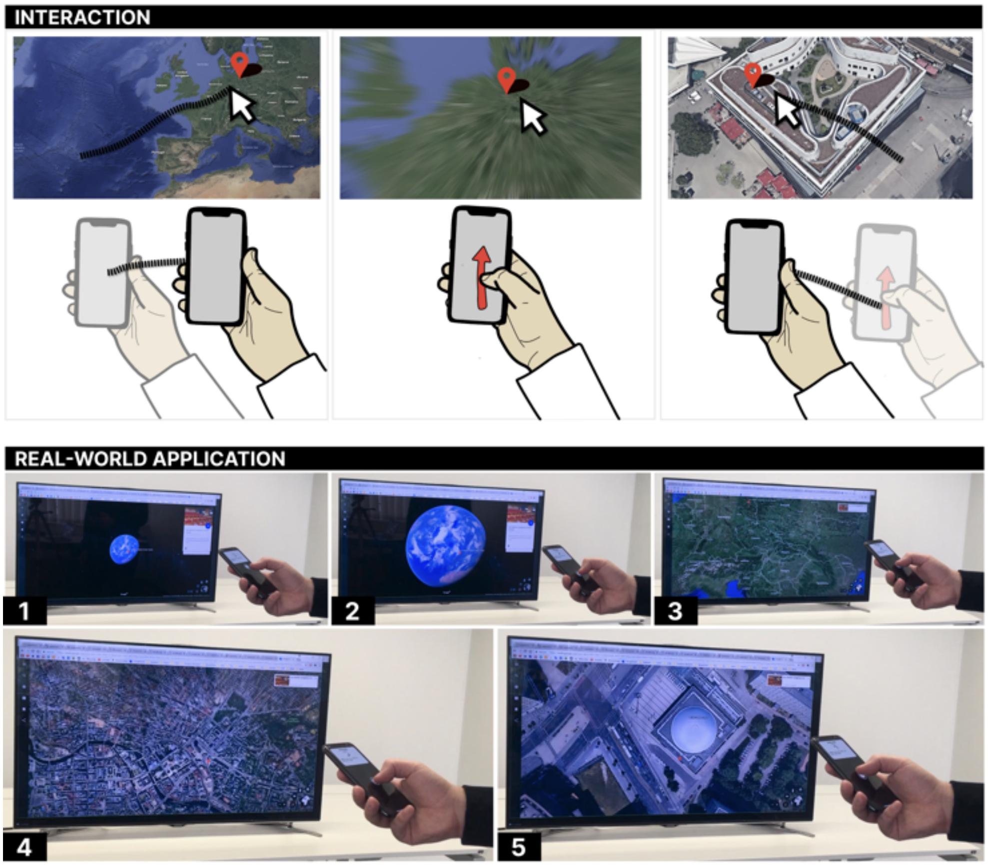


Figure 3.13: This figure demonstrates Pocket6 for map navigation. Users can control the 2D cursor by moving their smartphone in mid-air. Simultaneously, zooming functions are managed by swiping a finger up or down on the touchscreen. This dual-input approach, combining air gestures for cursor control and touchscreen gestures for zoom, results in a highly efficient and precise map navigation experience. 1-5 represents a sequence of the video captured in a real-world application.

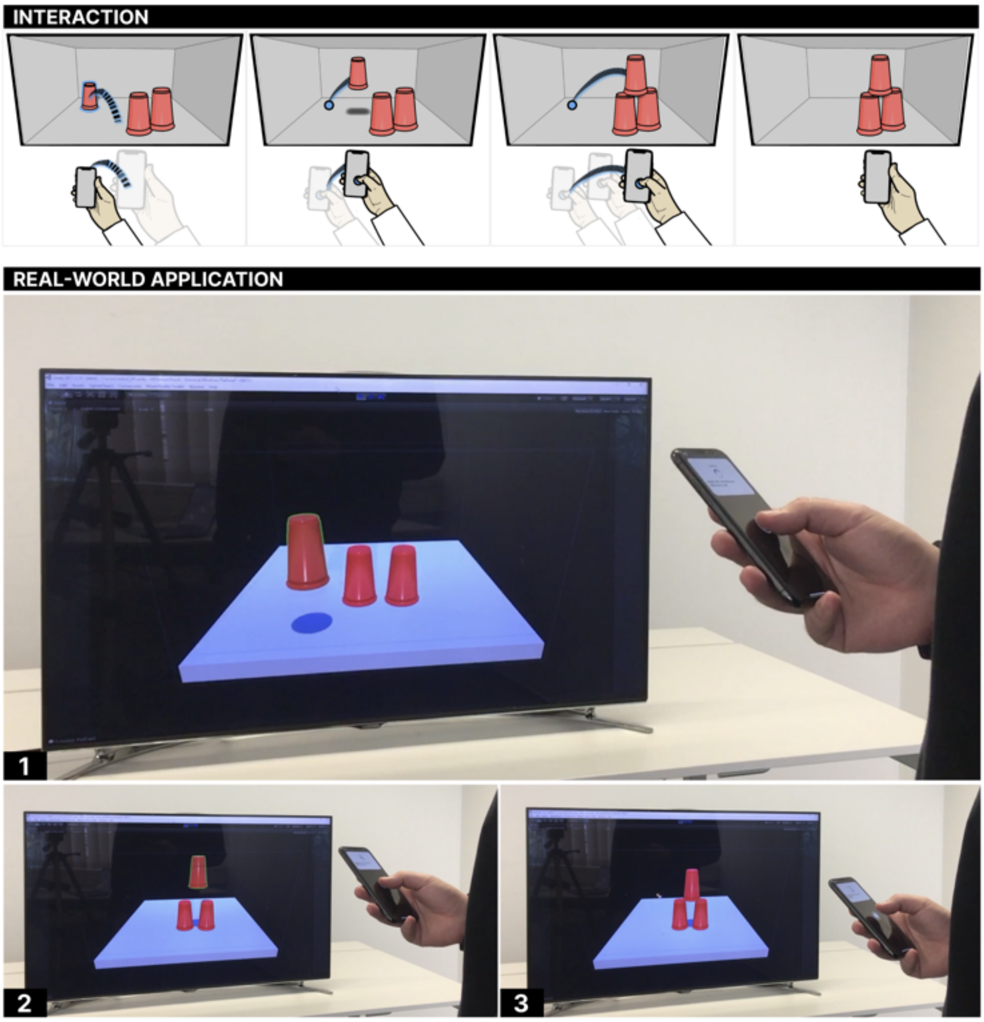


Figure 3.14: This figure demonstrates Pocket6 for basic 3D applications. It illustrates the use of Pocket6 in a game where players build a pyramid from cups. The process begins with the player tapping on the smartphone's touchscreen to grab a 3D cup. Following this, the player can move the smartphone in mid-air to translate the cup's position in the virtual space. 1-3 represents a sequence of the video captured in the real-world application.

Cups Game

Another possibility of combining mid-air gestures with touch events is seen in the game where players construct a pyramid from cups, as depicted in Figure 3.14. Players can grab 3D objects, such as a cup, by simply tapping on the smartphone's touchscreen. Subsequently, by maneuvering the smartphone through the air, the 3D object can be translated in a corresponding manner.

Select, Drag, Drop

We also see a lot of potential in simplifying the copy & paste and drag-and-drop task, which often causes problems with remote controllers, especially across different documents, as the selection with the touch gestures alone tends to often be challenging. In our approach, illustrated in Figure 3.15, this process merges touch inputs for text selection with mid-air movements for highlighting text and switching between documents. The paste function is activated simply by lifting a finger off the phone.

Interior Designer

Lastly, we demonstrate a more advanced 3D application that facilitates the translation and rotation of 3D objects. In the implemented interior designer app, illustrated in Figure 3.16, users can grab a 3D object, like a bed, by touching the smartphone's touchscreen. By moving the smartphone in mid-air, the 3D object can be translated accordingly, while they can simultaneously rotate the object by sliding their finger across the touchscreen from left to right. This integration of translation through smartphone movement with rotation via touchscreen sliding creates a highly effective input method, enabling users to rearrange 3D objects both efficiently and precisely.

After the formal experiments, participants had the opportunity to try out these applications and were asked to provide qualitative feedback. Participants tested all the applications using both the Pocket6 and VR controller setups. From their responses, it became evident that the participants were unhappy with the small size of the touchpads on the VR controllers, measuring just 1 inch in diameter. They expressed that due to the small touchpad area, the controller could not offer a fine degree of control, which made it difficult to perform swipe gestures. For example, in the furniture rearranging app, this limitation meant that users had to frequently release and re-position their thumbs to effectively rotate the 3D objects with swipe gestures. In contrast, the larger touchscreen of a smartphone, measuring 5.8 inches, was found to be significantly more useful in these situations.

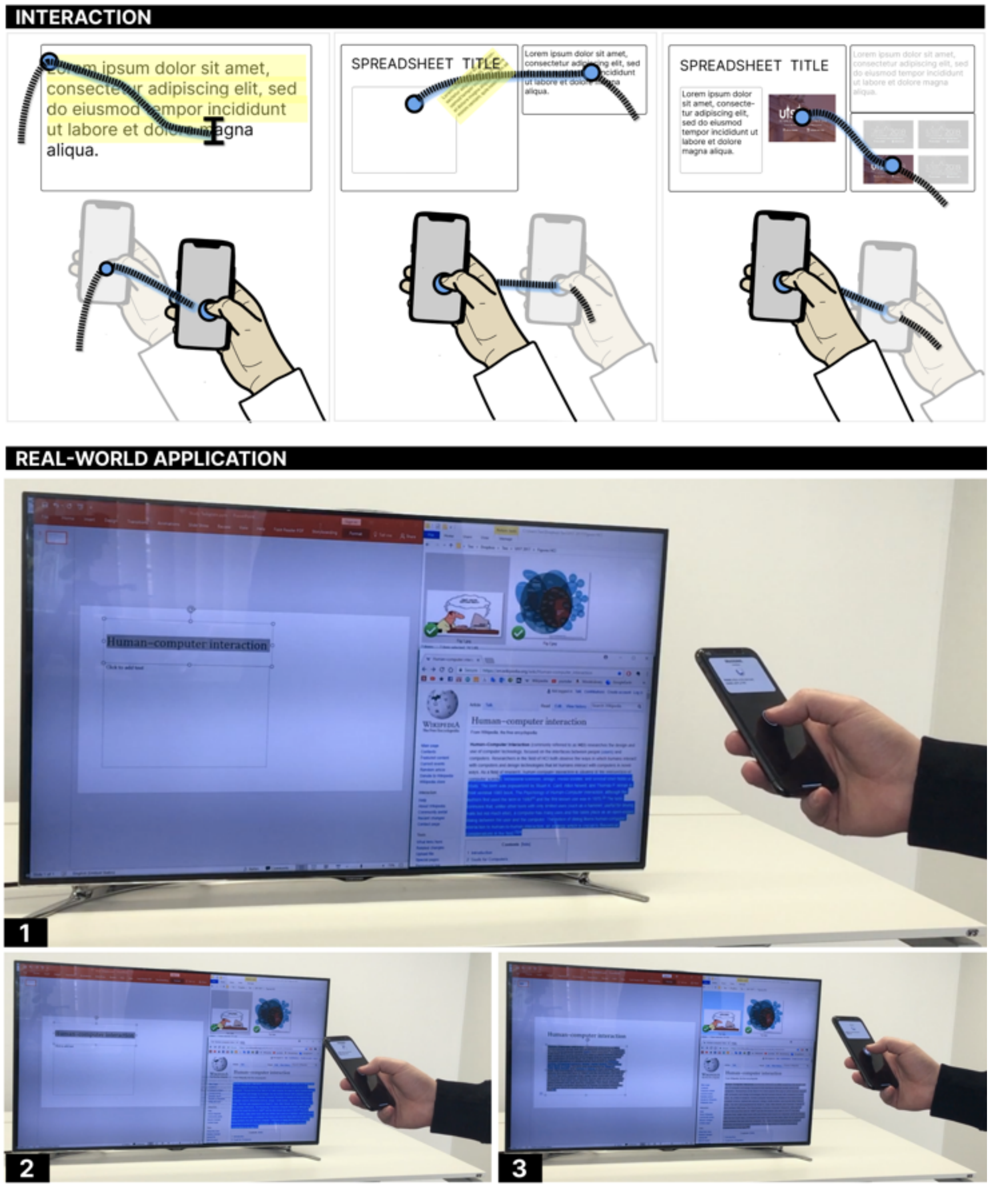


Figure 3.15: This figure demonstrates Pocket6's capabilities for text and 2D object manipulation. User controls the 2D cursor by physically moving their smartphone through mid-air. Simultaneously, users can perform touch-down and -up gestures to select a text or to drag-and-drop 2D objects, for example, for dragging and dropping selected text or figures into target applications such as a presentation or spreadsheet. 1-3 represents a sequence of the video captured in the real-world application.

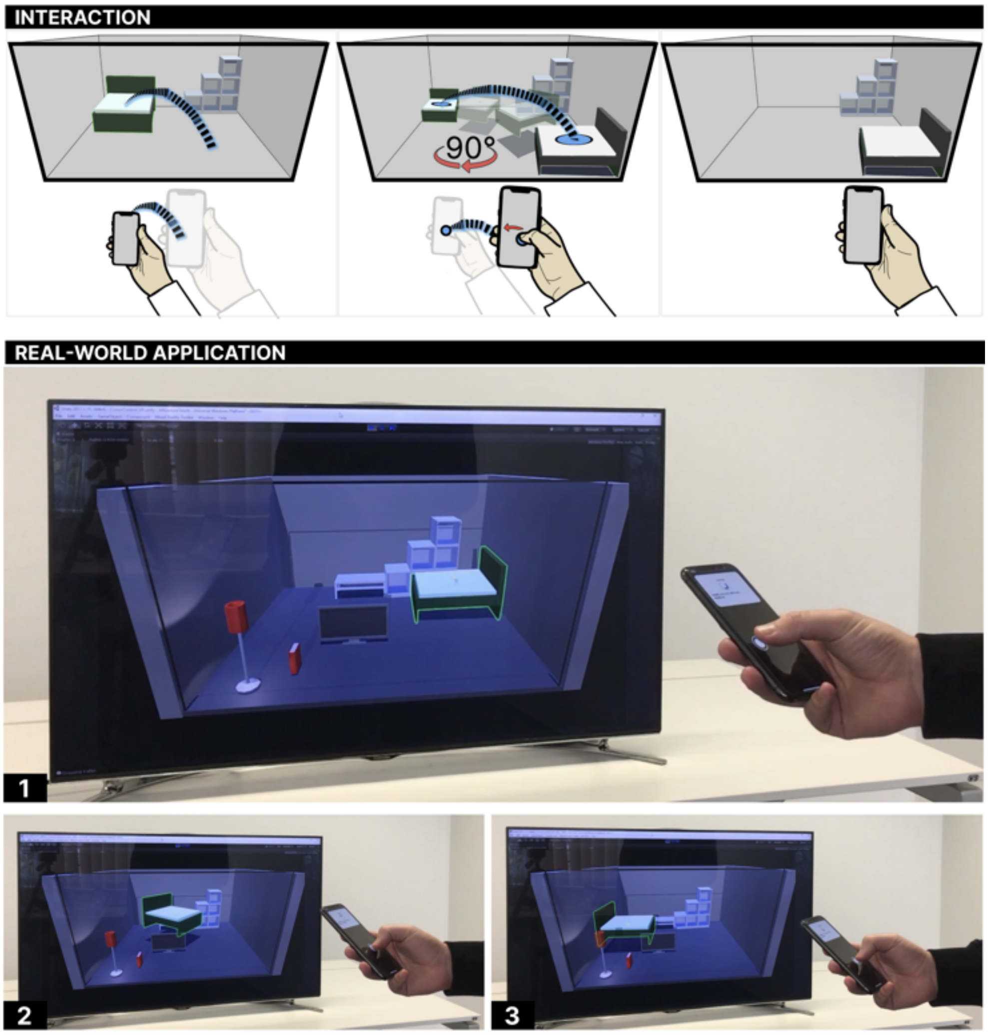


Figure 3.16: This figure demonstrates Pocket6's capabilities for 3D object manipulation. The integration of mid-air gesture input for object translation and touch input for rotation enables simple single-stroke 3D interactions. Illustrated here is the ability to manipulate 3D objects, like a bed, in an interior design app. Users grab objects with the smartphone's touchscreen and move them in mid-air for translation. Simultaneous rotation is achieved by sliding a finger across the touchscreen. This method combines ease and precision, enabling efficient and accurate rearrangement of 3D items. 1-3 represents a sequence of the video captured in the real-world application.

3.5 Summary

Across all experiments, we found that neither the performance nor the subjective opinions of the participants varied significantly between the Pocket6 *In-Front* and *Sitting* conditions. The *Hands-Down* condition was the least favored by participants, and they achieved notably lower accuracy with it than with other conditions. This outcome was unexpected, given our initial belief that the relaxed nature of the *Hands-Down* condition would make it the most comfortable for participants.

To summarize, our presented applications highlight Pocket6's practicality in various real-world scenarios and existing applications. We have shown how users can perform 2D interactions such as panning, drag-and-drop, and point-and-zoom. Additionally, we demonstrated the effectiveness of mid-air gestures for simple 2D selection tasks for text and objects, which are frequently challenging with only touch-based inputs. Moreover, we have demonstrated 3D interactions ranging from basic translation-based object manipulation to more complex 3D object handling, all achieved through subtle hand movements.

In the next chapter, we will continue with the exploration of our approach to utilizing a smartphone for spatial interaction with distant displays. We will explore the possibility of integrating additional input modalities, such as head and eye tracking, to enable spatial interaction beyond hand motion and touch inputs.

Parts of the following Chapter 4 have been published as:

Teo Babic, Harald Reiterer, and Michael Haller. 2020. Simo: Interactions with Distant Displays by Smartphones with Simultaneous Face and World Tracking. In: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems - CHI EA '20. Association for Computing Machinery, New York, NY, USA, 1–12. DOI: https://doi.org/10.1145/3334480.3382962

The responsibilities of this publication were as follows: I formulated the research questions, designed and implemented the research prototype, designed and conducted the user study, analyzed the study data, and wrote the paper. Harald Reiterer and Michael Haller supervised the work.

Supplemental Material

The QR codes below allow you to access the supplemental material by either scanning it using a mobile phone (print) or by clicking on it (digital).

Paper Video Figure Repository

4

Extending Smartphones' Spatial Tracking Capabilities by Simultaneous Tracking

Building upon the exploration of touch and hand movement inputs using a single device, as discussed in the previous chapter and related work [33, 172], many researchers proposed multimodal input modalities by using multiple tracking devices, such as combining smartphone touch with eye-gaze tracking glasses [250]. Implementing additional input modalities, such as tracking the head, eyes, and body, has allowed researchers to develop more efficient multimodal interaction techniques. However, on the negative side, these systems have also become more complex. Implementing inputs that relied on touch and the motion of multiple body parts (such as the hand [158], body [131], head [250] or eyes [250]) required augmentation of the user or the environment with additional tracking devices. While these multimodal interaction techniques have proven highly effective [259], minimizing the hardware complexity of these multi-device tracking systems remains a challenge. Currently, there is still no universally accessible single device that allows for straightforward spatial tracking of numerous body parts in the distant display context [39]. In this chapter, our primary objective is to implement a novel smartphone-based tracking approach that integrates touch input with motion tracking of the user's hand, head, body, and eye-gaze, all without the necessity for additional tracking devices. The particular problem we aim to solve for the distant display domain is the strong unavoidable bond between the number of enabled spatial input modalities and the hardware complexity. This dilemma has constrained distant display interaction design for years, hindering the deployment of in-lab systems for real-world testing and preventing end-users from adopting and learning novel interactions which have been proven beneficial many times in research setups (e.g. head- or gaze-pointing [250], 3D virtual hand [268]).

In this chapter, we present Simo, an innovative solution that transforms a regular smartphone into a spatial tracking device for user motion. It enables touch inputs and tracks the movements of the user's hand, head, body, and eye-gaze in the world-space, completely without the need for external tracking devices. The core concept of Simo is the simultaneous use of the smartphone's front and back cameras to enable concurrent face and world tracking, as depicted in Figure 4.1. Following this concept, we share the implementation and prototyping process behind Simo, which involved two connected smartphones combined within a specialized casing to enable the intended tracking capability. Lastly, we present two technical analyses that define the available interaction spaces and the tracking limitations.

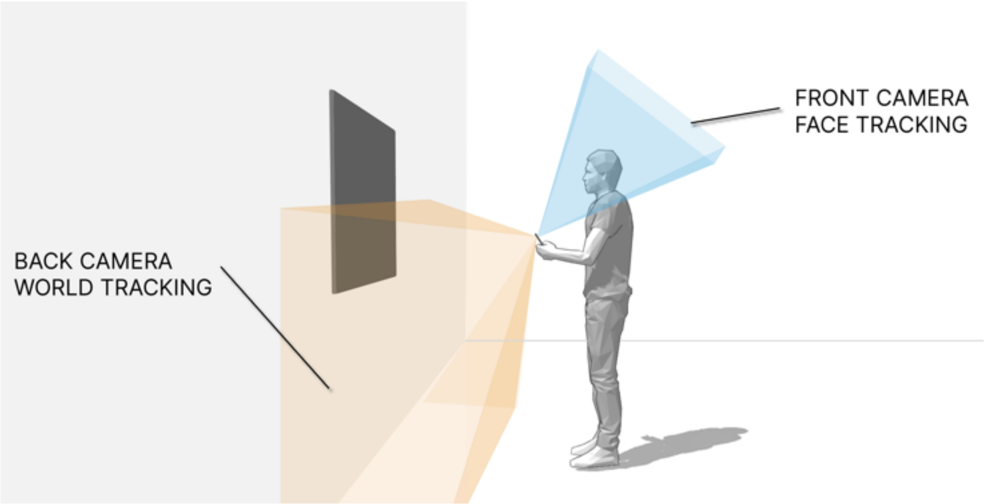


Figure 4.1: The Simo tracking principle uses simultaneously the smartphone's front- and back-facing cameras to enable spatial tracking of the user's hand, head, body, and eye-gaze for a distant display context.

4.1 The Simo Concept

Similar to Pocket6, also Simo's implementation concept draws from Apple's Augmented Reality toolkit, ARKit [6], that utilizes an inside-out mobile tracking method using the SLAM technology natively available in off-the-shelf smartphones. With the addition of a second phone, Simo extends the tracking framework of Pocket6 that enables touch and hand-motion inputs by the first phone's back camera for head, body, and eye-gaze tracking enabled by the second phone's front-facing camera tracking, as shown in Figure 4.1.

4.1.1 Prototype Versions

In the evolution of Simo, our development process was marked by the creation of various prototypes, each refined and updated in response to the latest iterations of iOS and ARKit. These versions ranged from initial rudimentary models and prototypes to more sophisticated designs, each leveraging advancements in smartphone technology and mobile augmented reality frameworks. This iterative process was crucial in adapting Simo's functionality to the evolving capabilities of mobile operating systems (e.g. iOS) and our understanding of the potential of ARKit and similar technologies.

Two-Phone Implementation

The initial implementation of Simo was based on Apple's iOS 13 and ARKit 2, which were limited to using only one specific camera at any given time. This version lacked the native support for simultaneous tracking using both cameras at the same time. Consequently, we decided for the addition of the second phone into the setup, enabling each device to utilize one of its cameras for tracking purposes. Inspired by the Modular Phone concept [228] and Magic Lenses [165, 173], we designed a smartphone device prototype. This prototype combined two smartphones within a single 3D-printed case, creating the impression for users of holding just one device. Figure 4.2 shows the prototype assembly and the final look. The lower phone (iPhone XS) was entirely covered by the top-mounted phone (iPhone XS Max), with the exception of a small notch to accommodate the front camera. The lower phone utilized its front-facing camera for face-tracking and transmitted this data via a local wireless network to the upper phone, which then merged it with the data from its own back camera performing world-tracking. With this prototype, we could start developing the Simo tracking principle, described in the following, without the native support of simultaneous tracking. In the subsequent sections, we dive deeper into the specifics of the tracking implementation.

Figure 4.2: Initial Simo prototype featuring a dual-smartphone setup encased in 3D-printed housing, cf. A. Component B employs the first phone performing world-tracking, while component C utilizes the second phone for face-tracking and subsequently transmits this data to the first phone for sensor data fusion.

Single-Phone Implementation

Throughout the development of this thesis and its associated project, we consistently updated and revised the initial Simo prototype mentioned above. A significant milestone in our update process occurred with the release of ARKit 3, which introduced native simultaneous tracking support on a single device. This breakthrough confirmed our initial development trajectory and thesis goals, as it theoretically allowed the Simo spatial tracking method to become accessible to everyone over the platform of regular smartphones. In the following chapters, we will dive deeper into each development phase, detailing the specific, step-by-step modifications implemented throughout the project's evolution. Finally, we conclude the final single device Simo implementation[1] in the Chapter Conclusion & Future Work of

---

[1] Github - Simo implementation: https://github.com/teobabic/simo

the thesis.

It is crucial to acknowledge that our implementation was based on the iPhone smartphone and the Apple iOS operating system. Consequently, we find it important to highlight the potential and limitations for implementing Simo on Android-operated smartphones, which hold over 70% of the global market share [247]. Currently, the Android operating system and its ARCore [76] framework for spatial tracking do not yet support simultaneous tracking. Therefore, for researchers focusing on the Android operating system, the two-phone implementation remains relevant as a workaround for in-lab exploratory and research purposes.

4.1.2 Tracking principle

In the following, we describe the Simo tracking principle, which consists out of three key measurements. Each measurement plays a crucial role in establishing a comprehensive tracking system for tracking multiple body parts and ensuring accurate real-world spatial mapping with the distant display:

- **Smartphone-to-Display** measurement: This establishes an absolute real-world spatial reference between the smartphone and the distant display.
- **Hand-to-Display, Head-to-Display, and Eye-to-Display** measurements: These track the user's hand holding the device, head pose (both position and orientation), and the pose of both eyes in absolute world-space relative to the distant display.
- **Body-to-Display** measurement: This approximates the user's body pose in absolute world-space relative to the distant display.

Smartphone- and Hand-to-Display Measurement

The upper smartphone's back camera performed world-tracking to determine the world-scale localization of the phone. This process allowed us to get the world-spatial pose of the smartphone, and, crucially, we could imply the pose of the user's hand holding the device. Additionally, we utilized ARKit's image-detection features ARKitWorldTracking [7] (or the Android equivalent ARCoreWorldTracking [77]) for localizing the distant display in world-space. For this, we displayed a small logo image on the distant display, which immediately disappears after it is located by the camera of the world-tracking phone, see Figure 4.3 A. Together, this is how we could get the pose of the smartphone and the real-world distant display in the same coordinate system of ARKit and provide the *Smartphone-to-Display* measurement. In terms of tracked body parts, this equals the *Hand-to-Display* measurement, which describes the spatial tracking between the device-holding hand in relation to the distant display. Following this, the smartphone was capable of adding additional body parts, that it could track, into the calculation of their spatial positions and orientations relative to both the smartphone and the distant display.

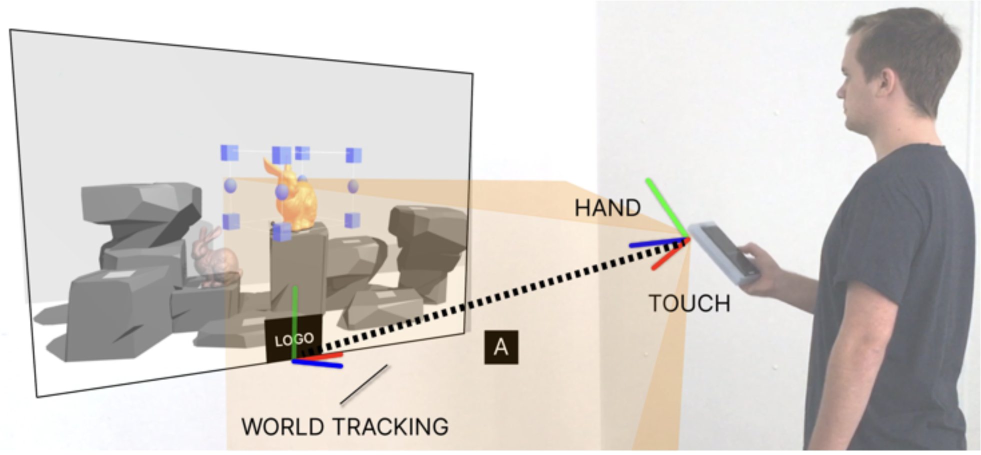


Figure 4.3: Illustration of the *Smartphone-to-Display* measurement, marked as A, which establishes a real-world spatial reference between the smartphone and the distant display. This is achieved by the smartphone’s back camera capturing a temporarily displayed logo on the distant display, enabling the spatial tracking of the hand holding the device in relation to the distant display.

Head- and Eye-to-Display Measurement

In the next step, we added the face-tracking capability of the second phone into Simo’s tracking framework. With the face-tracking, we could track the pose of the user’s head and both eyes. Consequently, the *Head-to-Display* measurement was calculated by merging the *Hand-to-Display* and the *Hand-to-Head* measurement, derived from the face-tracking camera of the smartphone, cf. Figure 4.4 B. This process was similarly applied to each eye, to determine the *Eye-to-Display* measurement, cf. Figure 4.4 C. All sensor data, including the information about the position and orientation of the head and eyes, was compiled on the first world-tracking phone. To achieve this, the second phone, dedicated to face-tracking, transmitted all relevant data to the first phone via the local wireless network.

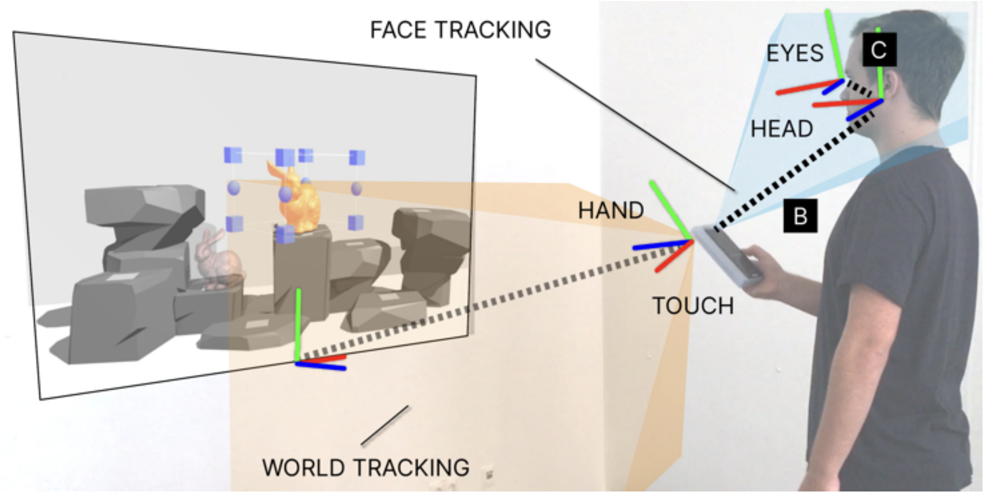


Figure 4.4: Illustration of the *Head-to-Display* and *Eye-to-Display* measurement, marked as B and C respectively. This enabled the spatial tracking of both the head and eyes relative to the distant display.

Body-To-Display Measurement

One of our key thesis objectives was to simultaneously track as many input modalities as possible to provide future developers with a rich set of input possibilities. In this process, we recognized that body tracking is a frequently requested feature, which typically requires complex camera-based tracking systems [113, 127]. Consequently, we developed a method that approximates the pose of the user's torso from the head and hand tracking. This method operates under the assumption that the body's position equals the head's position minus an y-axis offset (in our case 82.5 cm, the half of the average human height of 165 cm [167]), as depicted in the line marked with D in Figure 4.5. The body's orientation is also matched to the y-axis orientation of the head, resulting in the final *Body-to-Display* measurement.

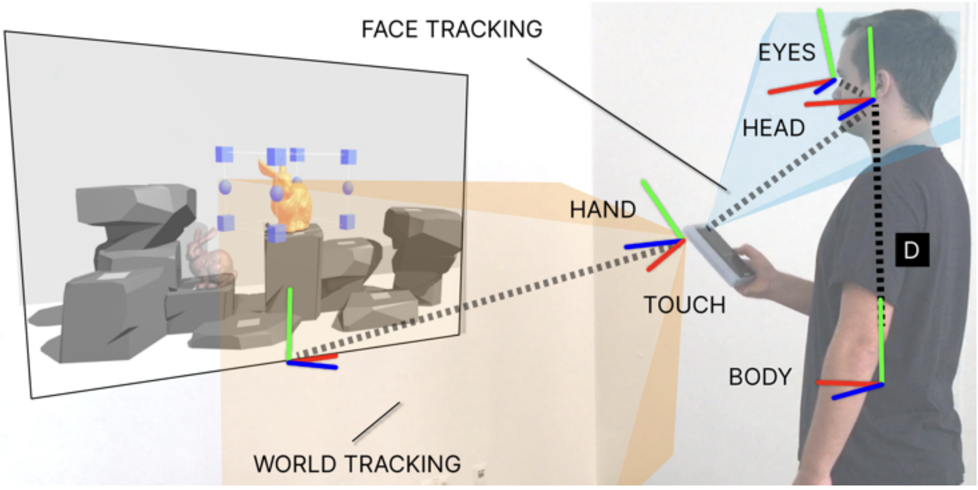


Figure 4.5: Illustration of the *Body-to-Display* measurement, marked as D. This enabled the spatial tracking of the user's body relative to the distant display.

Since the direct coupling of the body's and head's pose provides false-positive results, for example, if the user only turns the head, would the body also rotate (i.e. where head-only turns would incorrectly suggest body rotation), we implemented a high-pass filter. This filer checks whether the change in Euclidean distance between the hand and head poses exceeds a threshold k within a 20 ms window. This time window can be adjusted as needed in case other types of behavior are needed. When this condition is met, indicating simultaneous hand and head movement, we infer that the user is either turning or walking and accordingly update the body's position and orientation. A lower k value, cf. k=0.1 cm, resulted in more responsive but potentially less accurate body tracking, while a higher k value (e.g., 1.0 cm) led to less responsive but more stable tracking. In our exploration, we empirically set k to 1.05 cm. In the implementation, we have exposed the k threshold variable, allowing developers to adjust it for their specific use cases in which Simo is employed. The detailed algorithm for approximating the user's body pose is outlined in the subsequent code snippet[2]:

---

[2] Github - Simo algorithm implementation: https://github.com/teobabic/Simo/blob/master/Assets/SimoTracking/SimoTracking.cs

Listing 4.1: Code snippet defining the position and rotation of the user's body, based on the hand and head pose.

```
float k;                                                // Threshold
Vector3 hand, head, body;                               // Position
Vector3 previousHand, previousHead;                     // Position
Quaternion headRotation, bodyRotation;                  // Rotation
Vector3 headToBodyOffset = new Vector3(0, 82.5, 0);   // Head-to-body offset

void Update()
{

 ...

 // Execute every 20ms
 float deltaHand = Mathf.Abs(previousHand - hand);
 float deltaHead = Mathf.Abs(previousHead - head);

 bool isHandMoving = deltaHand > k;
 bool isHeadMoving = deltaHead > k;

 if (isHandMoving && isHeadMoving)
 {
    body = head - headToBodyOffset;                   // Body position
    bodyRotation.y = headRotation.y;                  // Body rotation
 }

 previousHand = hand;
 previousHead = head;

 ...

}

...
```

To summarize, Figure 4.6 shows all input modalities tracked by Simo. In the following chapter, we dive deeper into a detailed technical analysis of these tracking capabilities, defining the specific interaction spaces for each modality and highlighting their respective limitations.

Figure 4.6: In summary, Simo implements simultaneous use of the front and back camera of the smartphone for tracking of the absolute world-space pose of the user's device (hand), body, head, and eyes (gaze). Combined with touch inputs, this potentially enables, after a simple app download, any smartphone user to perform powerful multimodal spatial interactions for distant displays.

4.2 Analysis of Technical Capabilities and Limitations

To get a deeper understanding of Simo's advantages and limitations and to define its tracking range, we conducted three experiments. For the hand-tracking via the back camera, we did not see any particular limitations also based on our experience from the Pocket6 project, as long as we have enough light in the room and we do not have any obstacles, such as fingers or other users, blocking the camera's view for prolonged periods of time. Our primary focus, however, was on identifying the limitations related to head-tracking and understanding how smartphone movement affects it. We, therefore, conducted the following three experiments to address these issues.

4.2.1 Participants

A total of 20 volunteers (5 females and 15 males) were randomly chosen from a local organization to participate in our experiments, with an average age of $M = 28.1$ ($SD = 7.8$). All participants were regular computer and smartphone users, including two who were left-handed. The entire study, encompassing all three experiments, took approximately 15 minutes for each participant.

4.2.2 Analysis 1: Head Rotation Tracking

Head rotations typically range from $-80/80°$ left/right and $-60/50°$ down/up, as per the standard human range of motion [239], and are expected to be tracked within these limits. Consequently, our investigation focused on determining how far users can rotate their head until the head-tracking system cannot reliably detect the head rotation, leading to inaccuracies in the measured values.

Procedure and Task

In the experiment, participants were instructed to hold the smartphone approximately 40 cm in front of their eyes. For the baseline, a Vive tracker [265] was attached to the participant's head with a headband to accurately measure the head rotation. Throughout the experiment, head rotation data was simultaneously recorded by both the smartphone and the Vive tracker. Participants were instructed to rotate their heads once to the left, right, upward, and downward. They were advised to rotate their heads in a controlled manner to avoid injury and to stop the rotation at the point of discomfort, indicating their maximum range of motion. For the analysis of the results, we defined that the head-rotation tracking by Simo was deemed unreliable once the difference between the average head-rotation error, compared to Vive's measurements, exceeded $5°$.

Results

Figure 4.7 provides an overview of the smartphone's head-rotation-tracking performance. We found out that for the head rotation to the left and to the right, the smartphone's head tracking was less accurate once the head rotation exceeds 60°, see Figure 4.8 B. Similarly, the tracking accuracy reduced at 67° for upward rotations and at 35° for downward rotations, illustrated in Figure 4.8 A.

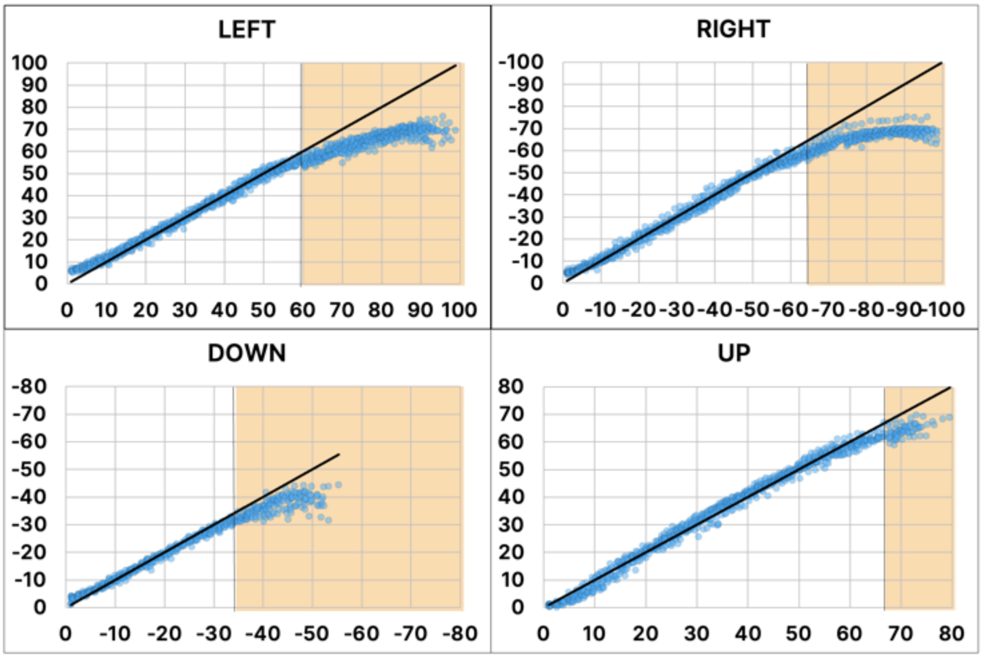


Figure 4.7: The figure displays all head rotation data collected through Simo (indicated by blue dots) against data from Vive (shown in black). The orange highlighted sections represent angles exceeding the reliable tracking limit, where head rotation tracking becomes inaccurate, with an angular error greater than 5°.

4.2.3 Analysis 2: Smartphone's Front Camera Field of View

The effectiveness of head tracking can also decrease when the user's face moves beyond the camera's field of view. Therefore, the following experiment aimed to precisely determine the front camera's field of view where head tracking still remains possible.

Procedure and Task

Similar to the previous setup, participants were asked to hold the smartphone positioned in front of them. Participants were asked to hold their head still while slowly rotating the smartphone to the left, then to the right, upwards, and downwards.

During this procedure, the rotation of the smartphone was measured, recording the angle at which head tracking failed.

Results

On average, head-tracking failed to function when the smartphone was rotated left for $15.5°$ ($SD = 0.8°$), right for $15.5°$ ($SD = 0.8°$, upwards for $23.5°$ ($SD = 0.7°$), and rotating the smartphone downwards for $18.4°$ ($SD = 1.3°$). These results define an interactivity frustum for head-tracking. The frustum's origin point aligns with the front camera of the smartphone, featuring a horizontal angle of $31.0°$ ($SD = 1.5°$) and a vertical angle of $41.9°$ ($SD = 1.7°$), as depicted in Figure 4.8 C-D.

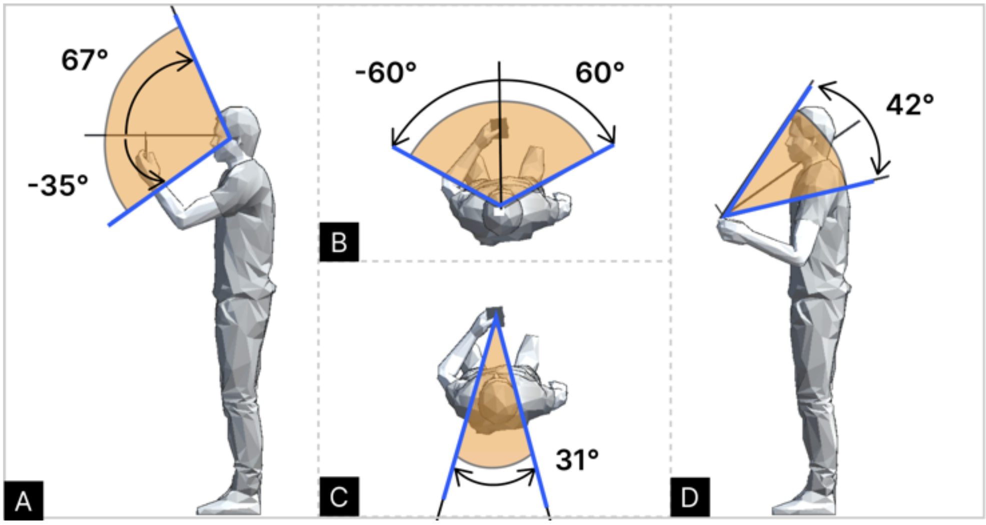


Figure 4.8: The diagram illustrates the range of head rotations for upward/downward A and leftward/rightward B movements, the horizontal C and vertical D field of view (FOV) for front camera-based face tracking.

4.2.4 Analysis 3: Smartphone's Holding Position

The last experiment focused on examining the changes in the smartphone's position and orientation when it is held in a typical manner, compared to its usage with Simo for head-pointing at a distant display.

Procedure and Task

In this experiment, participants were asked to look at the smartphone and play with it while holding it as they would normally. During this time, we captured the smartphone’s position and orientation relative to the user’s head. Later, we introduced a distant display (170 × 95.5 cm screen, 2 meters away from the user) with an interactive head-pointing cursor feature, as illustrated in Figure 4.9.

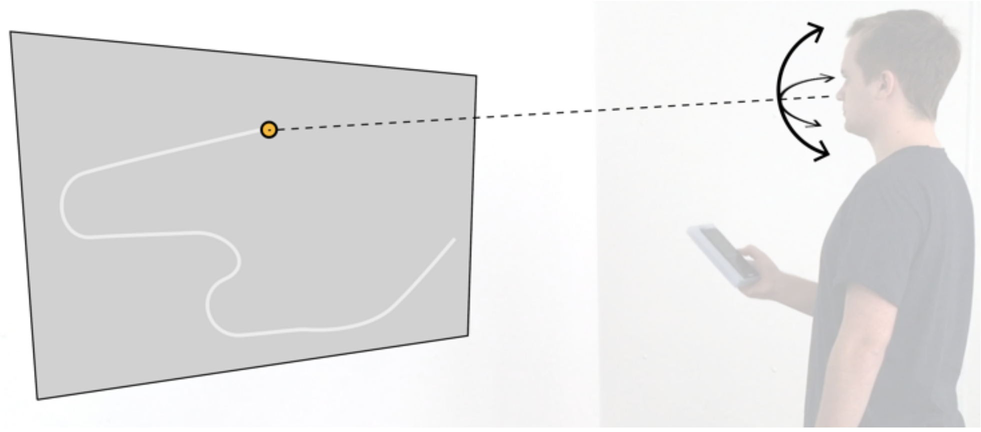


Figure 4.9: The setup of the third experiment, where participants interacted with a smartphone and a distant display. Participants were instructed to manipulate this cursor by using Simo with an implemented head-pointing interaction technique for two minutes.

Participants were instructed to interact with this display by moving a cursor remotely for about two minutes, after which they were asked to align the cursor at the center of the display. At this point, we captured the position and orientation of the smartphone again.

Results

The findings of the experiment indicated that, on average, participants held the smartphone at a $-50.1°$ angle ($SD = 2.0°$) along the x-axis relative to their head, as shown in Figure 4.10 E. During their interaction with Simo and the distant display, this angle shifted to $-55.0°$ ($SD = 2.9°$), as illustrated in Figure 4.10 F. This indicates that when users transitioned to using Simo, they slightly lowered their hand by approximately $5°$. In contrast, the angles on the z- and y-axes remained relatively constant, with changes less than $1°$.

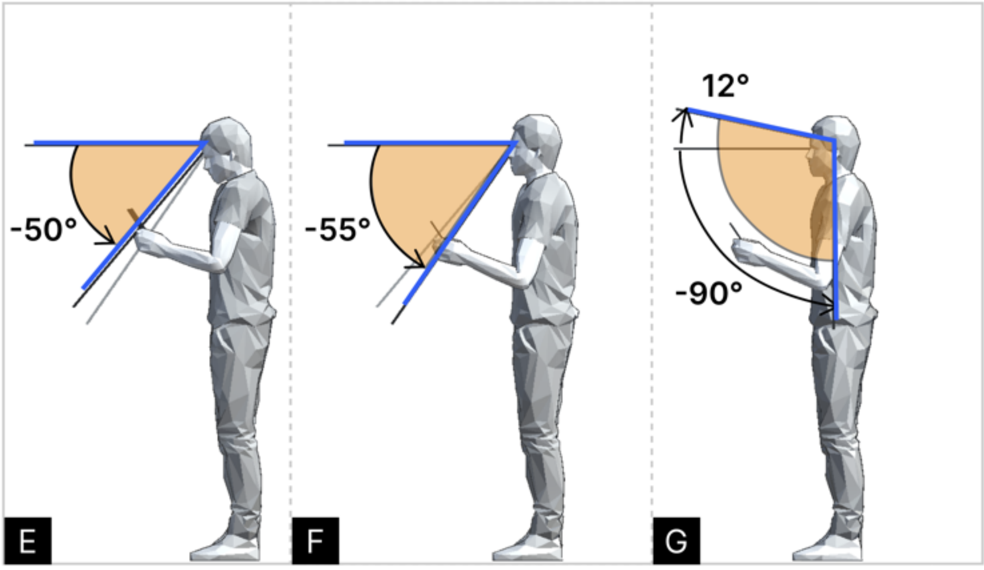


Figure 4.10: The diagram illustrates the typical angles of holding a smartphone in regular use E and while Simo is used F. Furthermore, is illustrates Simo's resulting interaction space G.

4.3 Resulting Interaction Space

To summarize, we draw from the results of all three experiments and outline the resulting interaction space of Simo, as shown in Figure 4.8 sections B-D and in Figure 4.10 G. The vertical angle of −55°, representing the typical smartphone holding angle during Simo interaction (refer to Figure 4.10 F), combined with the vertical head-rotation tracking limits of 67° upward and −35° downward (as shown in Figure 4.8 A), resulting in a vertical interaction span of 12° upward and −90° downward, as depicted in Figure 4.10 G. It's important to note that, within this range, the head-rotation tracking is highly reliable. Beyond the 12° limit, the tracked head-rotation is still possible. However, it will gradually become less accurate (resulting in "drift") until eventually becoming unreliable at angles exceeding 22.1° above the eye-line. The resulting horizontal angle of the Simo interaction space maintains a consistent angle of 60.0° on both sides, as illustrated in Figure 4.8 B.

4.4 Summary

In this chapter, we presented Simo's implementation process, from its initial two-phone prototype to its final single-phone implementation, both focusing on enabling simultaneous use of the smartphone's front and back cameras. We detail Simo's implementation, highlighting its capability to track hand, head, body, and eye-gaze movements in the world-space, without the need for external tracking devices. Finally, we presented a technical analysis that outlines the available interaction spaces and identifies the tracking limitations inherent to our smartphone-based tracking approach.

To summarize, the analysis provided a clear understanding of Simo's effective interaction space, considering both head-rotation and field-of-view tracking limitations. The conducted analyses highlighted specific angular limits for accurate tracking and exposed thresholds beyond which reduced precision occurs. Additionally, the analysis investigated smartphone holding angles, noting minor alterations during interaction with Simo.

In the next chapter, we delve further into the exploration of our approach for utilizing smartphone-based spatial interactions with distant displays. We show how we designed and evaluated an array of multimodal interaction techniques suitable for inside-out systems such as Simo. Furthermore, the following chapter presents the results of our usability evaluation. Here, the goal was to identify the most effective multimodal interaction technique for interacting in contexts ranging from simple 2D pointing to complex 3D object manipulation on a distant display using only smartphone-based tracking.

Parts of the following Chapter 5 have been published as:

Teo Babic, Harald Reiterer, and Michael Haller. 2022. Understanding and Creating Spatial Interactions with Distant Displays Enabled by Unmodified Off-The-Shelf Smartphones. In: Multimodal Technologies and Interaction. 2022; 6(10):94. DOI: https://doi.org/10.3390/mti6100094

The responsibilities of this publication were as follows: I formulated the research questions, analyzed the existing field of research, designed and created a taxonomy, designed and implemented the research prototype, designed and conducted the user study, analyzed the study data, and wrote the paper. Harald Reiterer and Michael Haller supervised the work.

Supplemental Material

The QR codes below allow you to access the supplemental material by either scanning it using a mobile phone (print) or by clicking on it (digital).

Paper

Repository

5

Designing and Evaluating Multimodal Interaction Techniques for 2D and 3D Distant Displays Interaction

In this chapter, we discuss the exploration of interaction techniques suitable for inside-out systems like Simo and address the challenge of selecting the most effective technique for a range of applications, from simple 2D pointing to complex 3D object manipulation. This selection process often involves trade-offs due to the limited input modalities enabled by inside-out systems [234]. Researchers frequently confront the dilemma of choosing between richer interaction possibilities that require external trackers and simpler setups that offer fewer modalities [200]. We conduct a detailed analysis of different interaction techniques, drawing inspiration from existing related work on touch [60], ray-casting [180], plane-casting [120], and virtual-hand interactions [12], where researchers highlight each techniques efficacy and limitations.

Based on our Simo prototype, using simultaneously the front and back camera of the smartphone, we present the design of multimodal interaction techniques that combine *touch* inputs with realtime world-space tracking of the user's *hand*, *body*, *head* motion, and *gaze*. Furthermore, we show how our systems can enable very precise spatial interactions through the integration of a dual-precision approach (i.e. input refinement mode), where a primary input mode is used for fast and coarse interactions while another secondary mode can be used for precise refinement. The main part of this chapter is represented by the two studies, evaluating our novel smartphone-based tracking approach and comparing the performance of our multimodal interaction techniques, inclusive of the refinement mode, within both 2D and 3D distant display contexts. Our aim in these studies was to identify the most effective methods for user interaction. With this exploration we evaluated the capabilities of the current system but also showed general future directions where the balance between interaction richness (i.e. amount of enabled modalities and interaction techniques), and system complexity can be optimized for better user experience and wider application [59].

Finally, we showcase how Simo can be used for a variety of 2D and 3D interaction scenarios, ranging from simple games to advanced 3D user interfaces. Additionally, we demonstrate the capability of our smartphone-based tracking system in enabling not just the input but also the output of 3D user interfaces, leading to enhanced rendering of 3D content on a 2D display.

5.1 Designing the Interaction Techniques

To evaluate the effectiveness of our novel smartphone-based tracking method, we designed interaction techniques for which we conducted a user study to gather quantitative performance data. While in our preliminary work with Simo, explained in Chapter 4, we performed a technical analysis of our system's tracking range, in this evaluation, we focused on analyzing user input performance with a variety of multimodal 2D and 3D interaction techniques. Based on prior research on smartphone-enabled interaction with distant displays, we found studies that compared various techniques: touch with head-pointing [234], touch with ray-casting [180, 234], touch with a virtual-hand [28, 107], and touch with gaze [259]. Consequently, we decided to design *touch*, *ray-casting*, *virtual-hand*, and *head-pointing* techniques for our comparative study. The primary goal was to find interaction techniques that require little physical overhead and which provide high accuracy, even while interacting with very small target objects. Due to the lacking implementation of eye-gaze tracking in ARKit, at the time as we conducted our study, we were unable to include eye-gaze techniques into this particular study. It is noteworthy that eye-tracking was subsequently added to ARKit, which we then included in our open-source Simo project[1].

---

[1] Github - Simo implementation: https://github.com/teobabic/simo

Similarly, as in other works, we included a dual-precision approach [138, 180], where one primary modality or input mode is used for coarse ("suggesting") and a second for fine-grained inputs ("refinement"). This approach was proven to be very useful for accurately selecting very small targets. Noting that most related studies focus on these techniques in the context of 2D user interfaces, we additionally incorporated a 3D user interface task to test their efficacy with a distant display.

In the following, we discuss the design of our four distinct interaction techniques, illustrated in Figure 5.1:

- **Touch** The *Touch* technique maps the user's finger movements on the smartphone's touchscreen into cursor movements on the distant display. The entire surface of the distant display (170 cm × 95.5 cm) is correspondingly mapped to the touchscreen area of the phone, encompassing its full width (6.9 cm or 1242 px) and the proportional height (3.9 cm or 698 px). In the primary mode, a 1:1 mapping ratio is employed, whereas the refinement mode utilizes a 2:1 mapping ratio, following the suggestion from Kytö et al. [138], in doubling the finger movement needed to move the cursor the same distance as in the primary mode.
- **Pointer** The *Pointer* technique employs ray-casting, projecting a ray forward from the top of the smartphone. In the primary mode, *Pointer* utilizes an absolute pointing method, while the refinement mode adopts a relative pointing approach with a 2:1 ratio. This means that starting with the current cursor position when switching to the refinement mode, only a half-degree per physically performed full-degree in smartphone rotation is applied to the ray's orientation update.
- **Hand** We implement a virtual-hand metaphor to map hand motions, where a 2D/3D virtual hand or cursor mimics the user's hand and directly follows its movement. The hand's position is tracked within a virtual control space measuring 50 × 20 × 20 cm, which aligns with the display space of 170 × 95.5 × 95.5 cm (1920 × 1080 × 1080 px) in the primary mode. Similar to earlier techniques, we shift to a 2:1 relative mapping in the refinement mode for more precise control.
- **Head** We also included head-pointing, a fundamental technique in augmented and virtual reality setups, yet often overlooked for distant display interactions. The *Head* technique works similarly to the *Pointer* technique,

with the difference that the head position is the origin of the ray, and the ray's direction is defined by the direction the head or nose faces. In refinement mode, *Head* technique mirrors the *Pointer* technique's functionality, where the degree applied to the cursor gets halved when switching to the refinement mode.

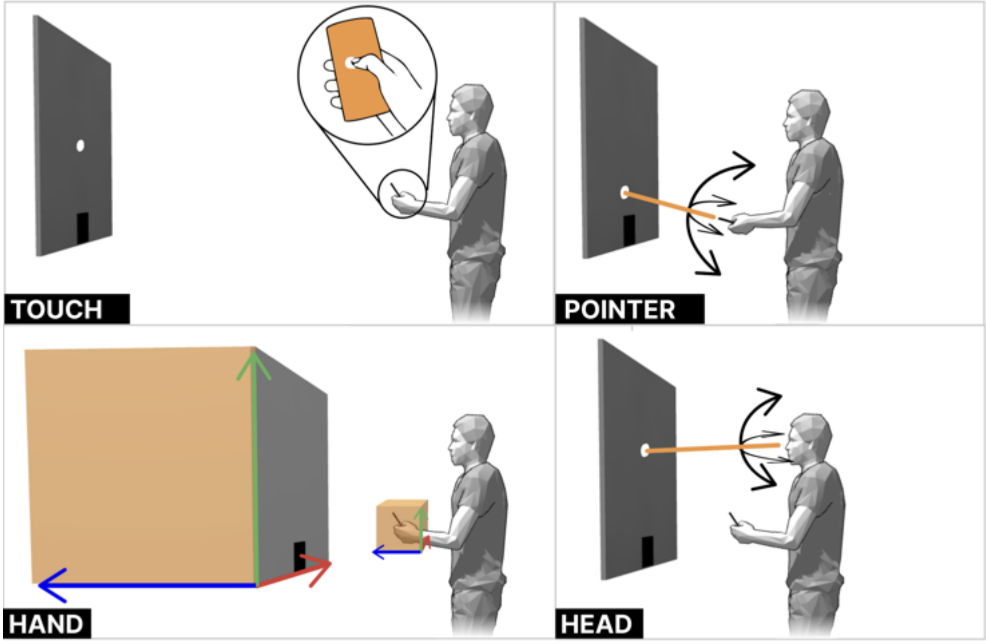


Figure 5.1: Multimodal interaction techniques enabled by Simo. The *Touch* technique maps finger movements on the smartphone's touchscreen to cursor movements on a distant display, with a 1:1 mapping ratio in primary mode and a 2:1 ratio in refinement mode for precise control. The *Pointer* technique uses ray-casting from the smartphone, offering absolute pointing in primary mode and halved cursor movement in refinement mode for enhanced accuracy. The *Hand* technique employs a virtual-hand metaphor, where a virtual hand or cursor in 2D/3D space mimics the user's hand movements, offering direct control within a defined virtual space. The *Head* technique incorporates head-pointing, where the head's position and orientation define the ray's origin and direction, with refinement mode reducing cursor movement degree for precise interactions.

5.1.1 3D Interaction

While all four interaction techniques can be used for 2D user interfaces, only *Hand* directly includes native 3D capabilities. The remaining three techniques can be adapted for 3D by using their standard 2D methods for cursor positioning on the x- and y-axes and employing the *Hand* technique's "fishing reel" metaphor for z-axis control [37]. In the *Pointer* technique, device tilting moves the cursor across the x/y plane, while moving the hand forward or backward adjusts its cursor's position along the z-axis. Likewise, with the *Head* technique, users direct the cursor on the x/y axes through head-pointing, and use hand movements back and forth to control the depth of the cursor on the z-axis.

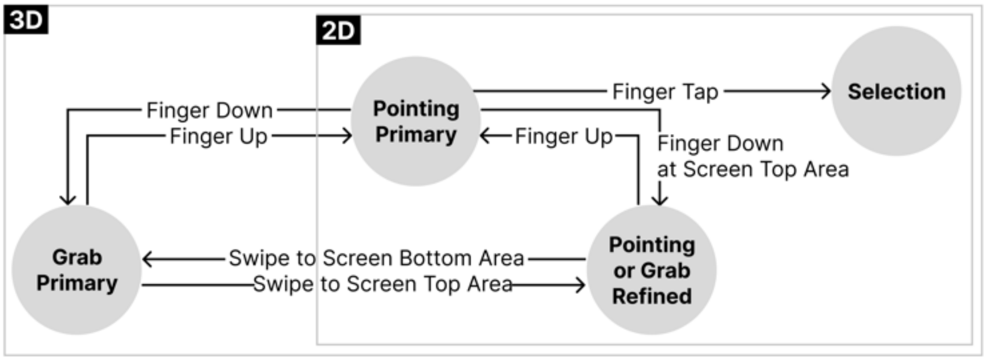


Figure 5.2: Overview of touch input states and transitions. It shows the logic behind the touchscreen's use for selecting 2D targets and manipulating 3D objects. It shows how the touchscreen is divided into two horizontal areas for different interactions: the bottom area for object interaction (selection and grab mechanics) and the upper part for activating refinement mode. The diagram also details the input state transitions, highlighting how a short tap (under 250 ms) triggers selection, while a longer touch activates the grab mechanic.

5.1.2 Touchscreen Interaction

For selecting 2D targets or grabbing 3D objects, the smartphone's touchscreen was used. The selection action was implemented as a short finger tap, triggered when the time difference between the touch down and up event was smaller than 250 ms, as in [234]. Touch was chosen as the primary selection method because of its deliberate interaction, along with its fast and reliable recognition. In addition to the tap, a grab mechanic was implemented used for 3D object manipulation, activated when the duration of the touch down event exceeded 250 ms. Since tapping was also the method to transition between primary and refinement techniques, the touchscreen was evenly split evenly into two horizontal areas, similar to the approach in the work of Nancel et al. [182]. The touchscreen's bottom area was dedicated to object interaction, while the upper part enabled the activation of the refinement mode. Figure 5.2 illustrates all touch input states and the transitions between them.

5.2 Evaluation

We conducted two studies aimed at comparing the precision and effectiveness of our multimodal interaction techniques when used with a distant display, in order to deepen our understanding of their practical application and performance. In the first study, we focused on 2D pointing and selection tasks, whereas the second study focused on 3D selection and manipulation tasks involving 6DOF docking tasks, as depicted in Figure 5.3. Across both studies, our objective was to identify not only the most effective primary and refinement techniques but also the optimal pairing of these techniques.

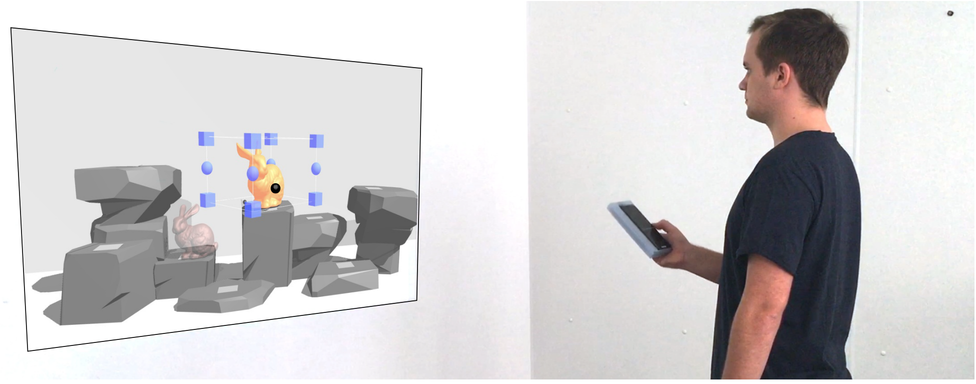


Figure 5.3: The figure depicts the apparatus used in our study, where the Simo smartphone app communicates wirelessly with a distant display and the connected computer.

5.2.1 Apparatus

In the overall setup, the Simo smartphone app transmitted all the tracked and processed user inputs to the distant display computer, which then rendered the user interface on the distant display. We used a projector (Epson EH-TW5650) with a projected image size of 170 × 95.5 cm (1920 × 1080 px) as a distant display. In contrast to the Pocket6 apparatus described in Chapter 3, we have transitioned from using a large display to a projector, solely because of the availability of hardware in our office. As in similar studies, participants were instructed to stand on a marked spot, centered and 2 m away from the projected image [180, 234]. The application for the study was developed in Unity using C# and was communicating with the smartphone via the local wireless network.

Signal Filtering

When performing simultaneous face- and world-tracking from a single handheld device, we need to be aware of an accumulated tracking noise error. This is because tracking noise in world-to-device camera tracking cumulatively adds up to the noise in device-to-user tracking. For instance, a jittery hand movement by the user impacts not only the pose of the smartphone but also the positioning of the head, body, and eyes. Therefore, we use the 1 € filter to minimize the overall tracking noise [44, 45]. The most effective filter parameters were identified as: for hand-position (mincutoff= 0.1, $\beta = 0.5$), for hand-orientation (mincutoff= 0.1, $\beta = 80$), for head-position (mincutoff= 0.001, $\beta = 0.01$), and for head-orientation (mincutoff= 0.1, $\beta = 25$). The frequency setting for all these filters was set at 60 Hz, with a dcutoff value of 1. The body pose calculation was performed post the signal filtering. Therefore, no additional filtering was necessary for body tracking.

5.2.2 Participants

A total of 18 paid volunteers (7 female, 11 male, average age $M = 28.9$, $SD = 6.0$) from various departments of the local organization participated in the study. Among them, 6 participants had prior intermediate experience with device ray-casting techniques, from using the Nintendo Wii controllers. All participants performed both studies consecutively within a total duration of one hour.

5.2.3 Study 1: 2D Pointing

In the first experiment, we investigated the performance of Simo and its four interaction techniques (*Touch, Pointer, Head, Hand*) in a 2D pointing task, while using them in combination with a distant display. We cross-combined all techniques with each other so that each possible combination of primary technique and optional refinement technique was considered. In this study, we used explicit target selection executed through a tap on the touchscreen.

Task

The task environment for our study was modeled after the experiment conducted by Vogel et al. [268]. In the task, participants needed to select a 5 cm diameter circular target using a cursor as shown in Figure 5.4. The cursor was represented as a 1 cm diameter circle, which changed into a 2 cm crosshair during the refinement mode. Targets were randomly placed throughout the entire distant display, maintaining a constant 50 cm distance from one target to the next. Before the study, we instructed participants to align the cursor with the target as precisely as possible, and to do so as quickly as possible. To discourage participants from excessive effort on accuracy over speed, we set a 5-second time limit for each trial, a duration determined from our pilot study results.

Design

The study utilized a repeated measures within-subject design, incorporating two variables:

- primary technique *(Touch, Pointer, Hand, Head)*, and
- refinement technique *(None, Touch, Pointer, Hand, Head)*.

We excluded the *Pointer+Head* combination from the study due to the incompatibility of these two techniques. Participants completed 20 target selection blocks per technique, with an additional 5 blocks reserved for training purposes. In total, each participant completed 380 trials (19 techniques × 20 blocks). The sequence of techniques was randomized for each participant to ensure variability.

For each trial, we measured the selection time and error, determined by the spatial difference between the cursor and the target position upon selection. After each technique, participants were asked to provide their subjective feedback by commenting on the currently experienced primary and refinement techniques regarding ease-of-use, physical, and mental demand. After the experiment, we asked participants also to select and comment on the three best and worst performing combinations. The subjective feedback sessions were also used as a break from the interaction to mitigate potential physical fatigue.

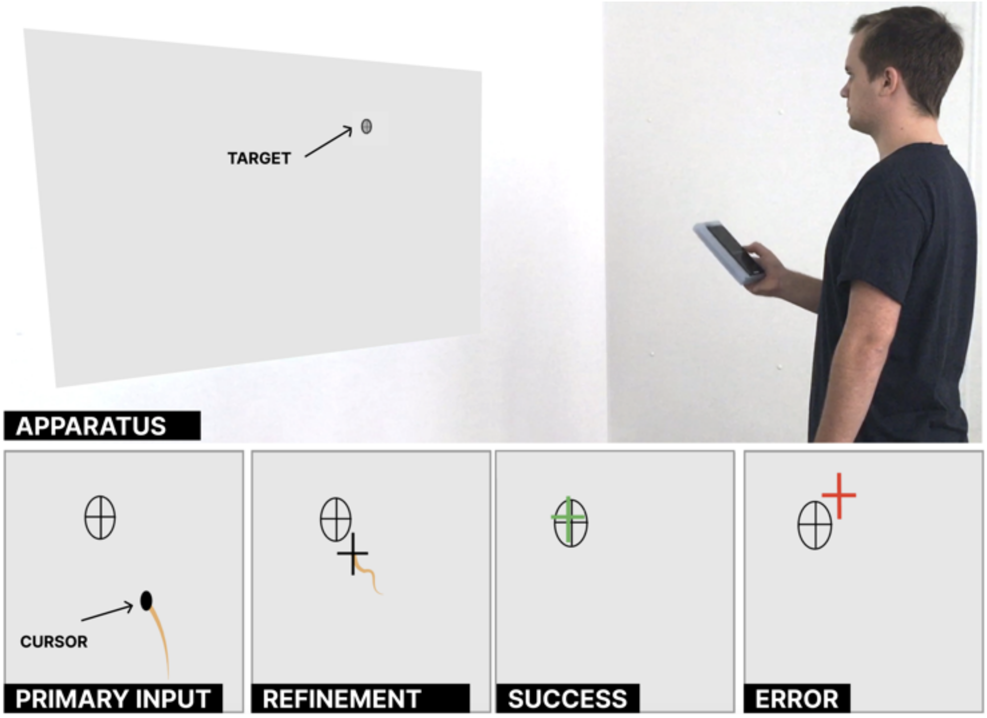


Figure 5.4: Figure showing the study setup, including both target and cursor visualizations, and also details the visualization variations in both primary and refinement modes.

Results

A repeated-measures ANOVA with an alpha level of 0.05 was conducted to analyze time and error data. When the assumption of sphericity was violated, indicated by using Mauchly's test, we used the Greenhouse-Geisser corrected values in the analysis. The post-hoc tests were conducted using pairwise t-tests with Bonferroni corrections. The analysis of time and error metrics included only successful target selections, accounting for 6782 out of a total of 6840 trials, or 99.2%. For the analysis, we considered all instances where a particular technique occurred as the primary mode or refinement mode respectively. It should be noted that all pairwise comparisons presented below adhere to a significance level of $p < 0.001$, unless noted differently.

Time We found a significant main effect on time for the primary modes ($F_{3,957} = 362.39, p < 0.001$), refinement modes ($F_{4,1276} = 174.30, p < 0.001$), and the interaction between the primary × refinement techniques ($F_{12,3828} = 93.19, p < 0.001$), as depicted in Figure 5.5

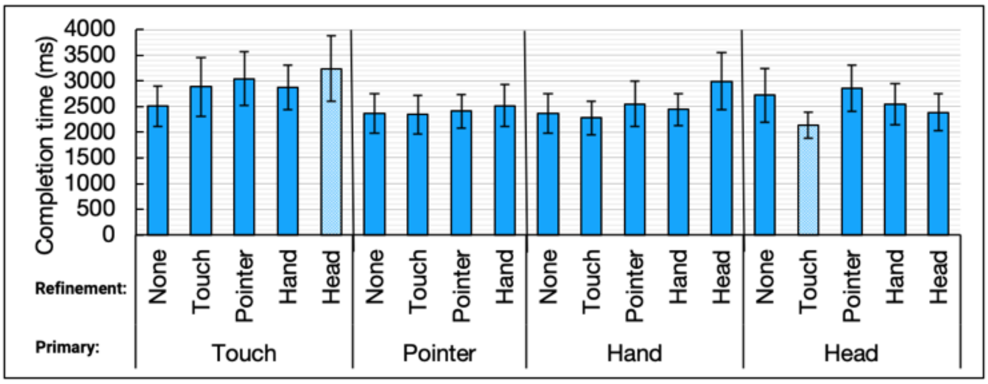
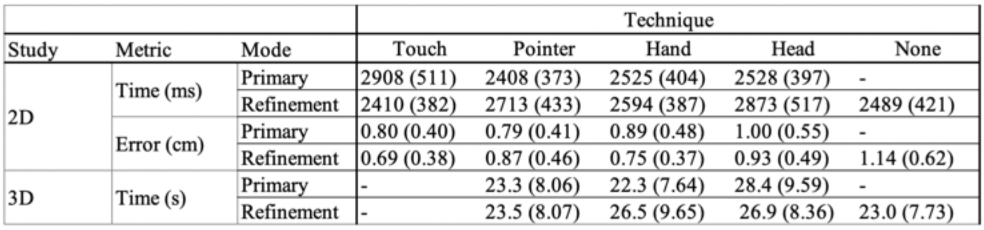
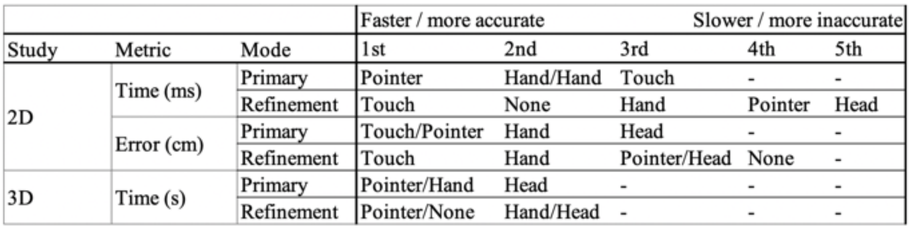


Figure 5.5: Mean selection times of our interaction techniques, with error bars representing the standard deviations and the pattern-filled bars highlighting the fastest and slowest technique.



			Technique				
Study	Metric	Mode	Touch	Pointer	Hand	Head	None
2D	Time (ms)	Primary	2908 (511)	2408 (373)	2525 (404)	2528 (397)	-
		Refinement	2410 (382)	2713 (433)	2594 (387)	2873 (517)	2489 (421)
	Error (cm)	Primary	0.80 (0.40)	0.79 (0.41)	0.89 (0.48)	1.00 (0.55)	-
		Refinement	0.69 (0.38)	0.87 (0.46)	0.75 (0.37)	0.93 (0.49)	1.14 (0.62)
3D	Time (s)	Primary	-	23.3 (8.06)	22.3 (7.64)	28.4 (9.59)	-
		Refinement	-	23.5 (8.07)	26.5 (9.65)	26.9 (8.36)	23.0 (7.73)

Figure 5.6: The study results contain the mean times and errors of our interaction techniques in both primary and refinement modes, with standard deviations enclosed in brackets. These numbers aggregate all attempts across all participants.



			Faster / more accurate				Slower / more inaccurate
Study	Metric	Mode	1st	2nd	3rd	4th	5th
2D	Time (ms)	Primary	Pointer	Hand/Hand	Touch	-	-
		Refinement	Touch	None	Hand	Pointer	Head
	Error (cm)	Primary	Touch/Pointer	Hand	Head	-	-
		Refinement	Touch	Hand	Pointer/Head	None	-
3D	Time (s)	Primary	Pointer/Hand	Head	-	-	-
		Refinement	Pointer/None	Hand/Head	-	-	-

Figure 5.7: The study results represent a ranking order, starting from the fastest and most accurate to the slowest and least accurate among our interaction techniques, covering both primary and refinement modes.

For the primary mode, which supported participants by moving the cursor quickly over a large distance, we found that the *Pointer* was the fastest technique, followed by *Hand* and *Head*, which were equally fast, while lastly *Touch* was the slowest, as shown in Figure 5.6 and Figure 5.7. For the refinement mode, which is used to refine the cursor's final position *Touch* was the fastest technique, followed by *None, Hand, Pointer*, and *Head*, which were all significantly different from each other. These results reveal that certain interaction techniques are great for fast and coarse pointing but are too slow for fine adjustments. For instance, while *Pointer* stands out as the quickest in the primary technique, it ranks as the second slowest in refining. Conversely, *Touch*, is the slowest for coarse pointing and proves to be the fastest for refining.

Upon evaluating the 19 different combinations of primary and refinement techniques, it became evident that some pairings performed better or worse than what their individual mode averages suggested. Our initial expectation was that the *Pointer-Touch* combination would emerge as the fastest overall, since it is a combination of the fastest primary and the fastest refinement technique. However, our results contradicted this assumption. Surprisingly, the *Head-Touch* ranked as the fastest technique, as shown in Figure 5.5. This underscores the significant gains in performance that can be achieved through the optimal understanding and pairing of techniques. Analyzing other pairings where *Head* functioned as the primary technique, like *Head-Touch*, *Head-Pointer*, and *Head-None*, we noticed their speed was only mediocre. However, the particular *Head-Touch* combination distinctly excelled, proving to be the overall fastest. On the other end of the scale, regarding the slowest techniques, the *Touch-Head* technique was the slowest, reinforcing the notion that individual technique's performances are worth considering, as *Touch* and *Head* are indeed the slowest in their respective primary and refinement categories. All significant differences in these combinations and their effects on performance are depicted in Figure 5.8.

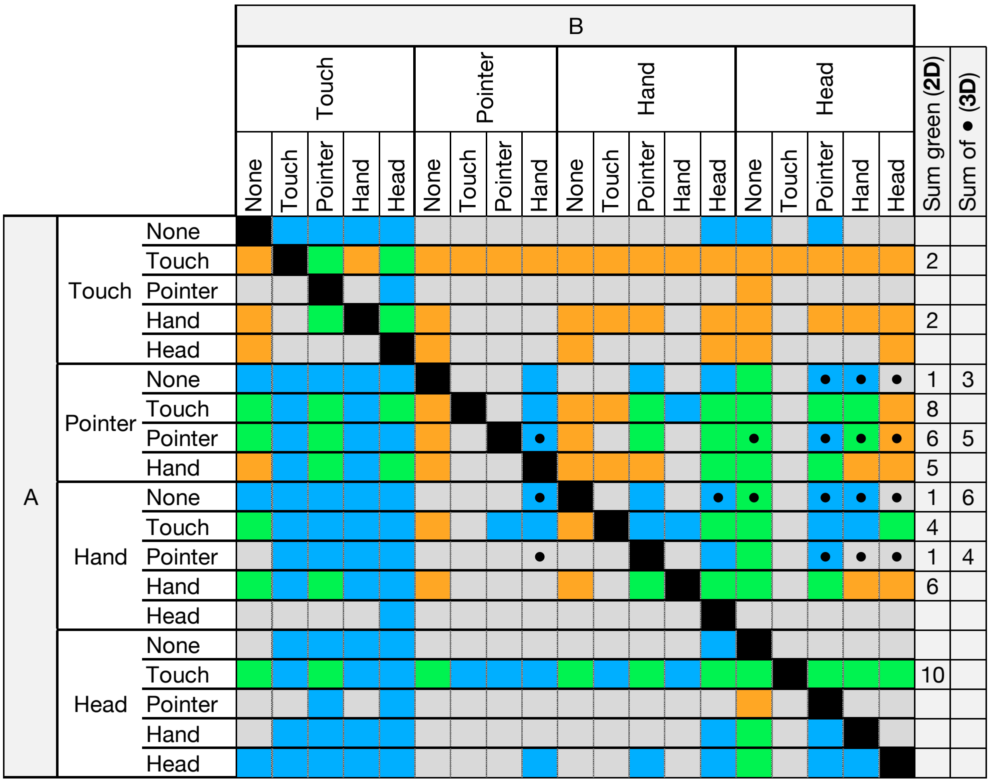


Figure 5.8: Matrix showing significant ($p < .05$) interactions among the 19 evaluated 2D techniques. In row A, technique x (blue for faster, orange for more accurate, green for both faster and more accurate) compares with technique y in column B. Additionally, the matrix shows all significant ($p < .05$) interactions among the 11 tested 3D techniques regarding time, represented by black dots (where ● indicates faster performance).

Error Significant main effects were observed in error rates between the primary techniques ($F_{3,942} = 65.249, p < 0.001$), refinement techniques ($F_{4,1256} = 160.278, p < 0.001$), and their interactions between the primary × refinement techniques ($F_{12,3768} = 20.540, p < 0.001$), as depicted in Figure 5.9.

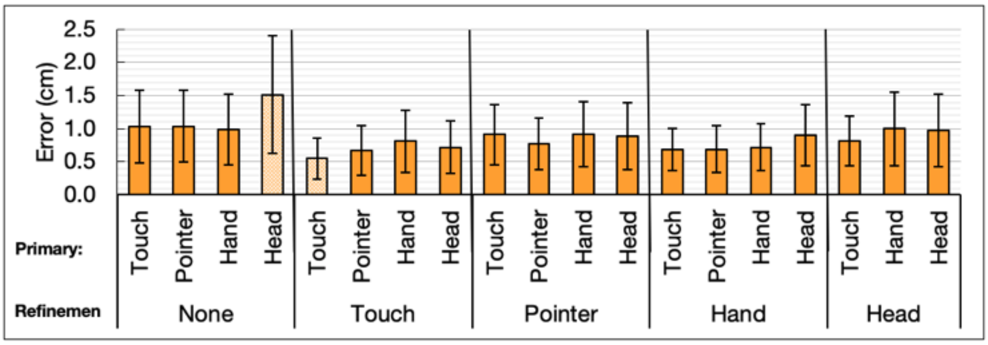


Figure 5.9: Mean pointing error across our interaction techniques. The error bars represent standard deviations, while the pattern-filled bars highlight the techniques with the highest and lowest accuracy.

In terms of accuracy, we need to point out that we were surprised by the overall accuracy of Simo. Regardless of the technique used, participants consistently achieved a precision level under 1.14 cm, as shown in Figure 5.6.

The study revealed that both *Touch* and *Pointer* were the most precise primary modes ($p = 0.57$), followed by *Hand* and *Head*. It was expected that the refinement mode was more important for accuracy since it is used to correct the pointing error. The results indicate that *Touch* emerged as the most precise refinement technique, closely followed by *Hand, Pointer*, and *Head*, which did not significantly differ in accuracy, and *None* ranking as the least accurate. From all 19 interaction techniques, the results show that techniques where *Touch* is used for refinement are the most accurate, with *Touch-Touch* standing out as the most accurate of all.

Our study indicates that a refinement mode can improve the selection accuracy for up to three times. As an example, we show in Figure 5.10 the comparison between the *Head-None* technique (without refinement) and *Head-Touch* technique using *Touch* for refinement.

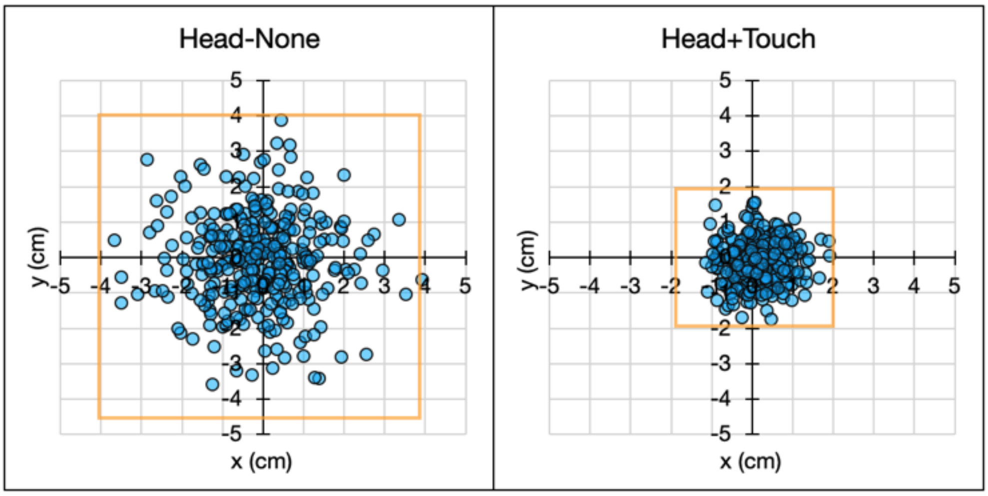


Figure 5.10: All selection points, showing the selection accuracy of the *Head-None* (no refinement) and *Head-Touch* (with refinement) technique. The coordinate system represents the distance from the exact target centre.

Our results align with the findings from Kytö et al. [138], who showed that refinement in head-mounted AR can improve accuracy for nearly five times, as well as the findings from Šparkov et al. [243], Chatterjee et al. [47] and Jalaliniya et al. [110], who found a threefold accuracy improvement for combining eye-gaze with hand gestures or head-pointing in distant display interaction.

Our findings reveal that even in interactive systems where only one input modality is provided, offering a refinement mode based on this same modality can significantly improve accuracy. This is evident from the better performance results of *Head-Head*, *Touch-Touch*, *Pointer-Pointer*, *Hand-Hand*, and *Head-Head* techniques compared to their respective *-None counterparts. Furthermore, in multimodal systems, mixing different modalities can lead to even greater accuracy improvements, as indicated by the performance comparison between *Head-Head* (uni-modal) and *Head-Touch* (multimodal). These observations align with the research of Kytö et al. [138], who also concluded that their unimodal head-head combination outperformed other primary techniques in terms of pointing accuracy in AR glasses setting.

Subjective feedback All 18 study participants expressed that the combinations *Head-Touch*, *Pointer-Touch*, and *Pointer-Pointer* are the best. This positive feedback was supported by comments, such as:

> *"Head-Touch, this is really good, since I only need to look at the target and the cursor is already there, then I just fine-tune. For bigger targets, you wouldn't even need a cursor." - P10*

> *"Pointer-Touch, this combination is really good; it very nicely calms down the cursor." - P6*

All 18 participants subjectively ranked the two combinations *Head-Touch* and *Pointer-Touch* among the top three preferred techniques. The *Head-Touch* combination was ranked first by 7 participants, while *Pointer-Touch* was chosen 4 times.

Furthermore, *Pointer-Pointer* was ranked in the top three by 6 participants (3 times as overall best), with arguments like:

> *"Pointer-Pointer, I like this one, since you do not need to do any touch actions, besides the tap." - P15.*

Moreover, the *Head-Head* technique was in the top three 3 times (once as the best), and the *Hand-Hand*, *Hand-Touch*, and *Touch-Touch* combinations each appeared 2 times in the top three, with each being rated as the best once.

Participants further provided valuable comments explaining, that touch-based techniques can feel slower compared to non-touch techniques, as explained in the following comment:

> *"Touch-Touch it's very precise, however, I need to perform multiple swipes, which makes me slow." - P16.*

Participants also pointed out that switching between modalities, for refinement, can reduce performance and be mentally demanding:

> *"Pointer-Hand, the input movement feels the same as Pointer-Pointer, although its harder. I would rather stay in the same modality as Pointer-Pointer." - P16*

> *"Head-Hand, I'm slow, since I need to stop rotating my head and start moving my hand. This makes me slow, and I need to actively think about doing it also." - P13*

> *"Head-Pointer, this is too complex, a bad combination." - P1.*

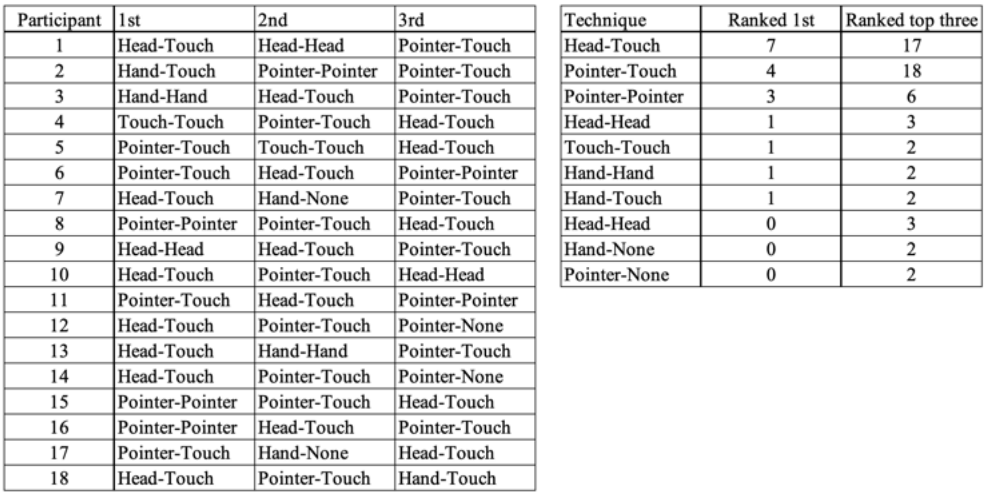
Finally, many participants were pleasantly surprised by the precision of solely using unimodal primary techniques without refinement:

> *"Hand-None, this is surprisingly precise." - P17*

However, they also saw the value in refinement mode, as explained in comments such as:

> *"Pointer-None, is fast and kind-of precise but I would still like to have a refinement mode." - P12*

In Figure 5.11, we show the subjective user rankings across all participants and all interaction techniques.



Participant	1st	2nd	3rd
1	Head-Touch	Head-Head	Pointer-Touch
2	Hand-Touch	Pointer-Pointer	Pointer-Touch
3	Hand-Hand	Head-Touch	Pointer-Touch
4	Touch-Touch	Pointer-Touch	Head-Touch
5	Pointer-Touch	Touch-Touch	Head-Touch
6	Pointer-Touch	Head-Touch	Pointer-Pointer
7	Head-Touch	Hand-None	Pointer-Touch
8	Pointer-Pointer	Pointer-Touch	Head-Touch
9	Head-Head	Head-Touch	Pointer-Touch
10	Head-Touch	Pointer-Touch	Head-Head
11	Pointer-Touch	Head-Touch	Pointer-Pointer
12	Head-Touch	Pointer-Touch	Pointer-None
13	Head-Touch	Hand-Hand	Pointer-Touch
14	Head-Touch	Pointer-Touch	Pointer-None
15	Pointer-Pointer	Pointer-Touch	Head-Touch
16	Pointer-Pointer	Head-Touch	Pointer-Touch
17	Pointer-Touch	Hand-None	Head-Touch
18	Head-Touch	Pointer-Touch	Hand-Touch

Technique	Ranked 1st	Ranked top three
Head-Touch	7	17
Pointer-Touch	4	18
Pointer-Pointer	3	6
Head-Head	1	3
Touch-Touch	1	2
Hand-Hand	1	2
Hand-Touch	1	2
Head-Head	0	3
Hand-None	0	2
Pointer-None	0	2

Figure 5.11: The results of the subjective user rating. The left table shows individual participant ratings for interaction techniques, ranked from 1st to 3rd. The right table aggregates these ratings, indicating how often each technique was ranked 1st the number of times it appeared in the top three choices across all participants.

5.2.4 Study 2: 3D Selection and manipulation

In our second study, we investigated the performance of Simo and our interaction techniques (*Pointer, Head, Hand*) while using them in a 3D docking task in combination with a distant display. We excluded the *Touch* techniques from this study, since the touch interaction conflicts with the 3D manipulation task, where touch is required for "grabbing" 3D objects. Similar to the previous study, we combined each primary technique with all refinement techniques, resulting in a total of 11 distinct interaction techniques.

Task

The design of our study is primarily based on the methodology used by Berge et al. [28] and is similar to other comparable experiments [31, 297]. Participants were required to perform a 6DOF docking task, shown in Figure 5.12 1, where they had to select a 3D object within a three-dimensional space and accurately align it with a target object, ensuring a match in position, rotation, and scale.

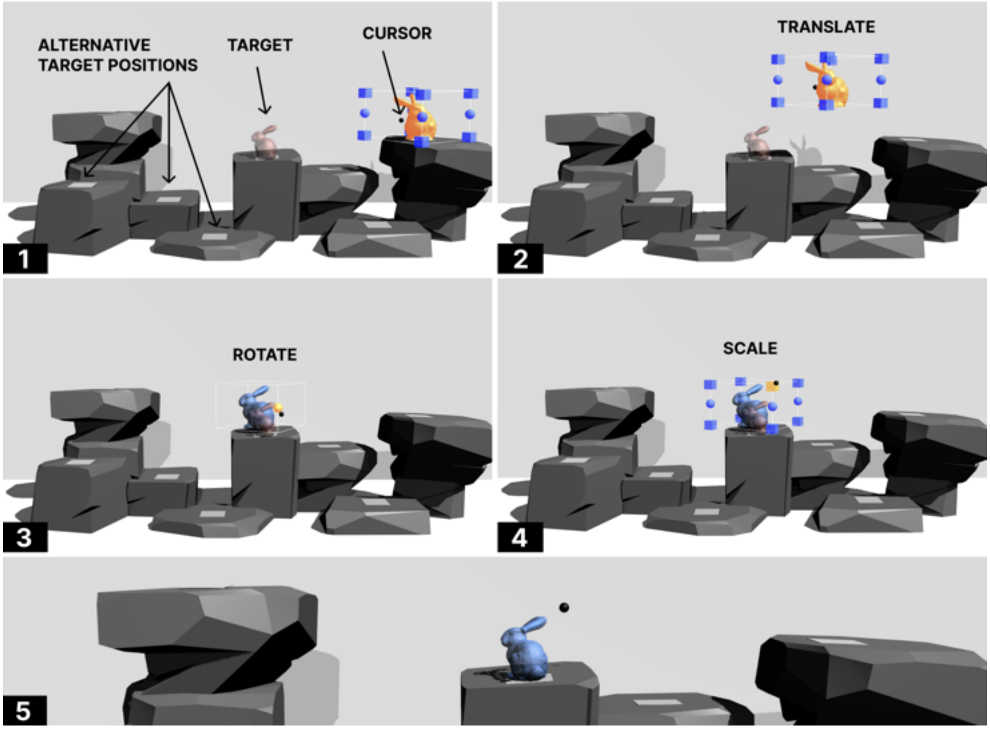


Figure 5.12: The depiction of the implemented 3D docking task: 1 shows the target, cursor, and alternate target positions. 2-4 displays the progression of the docking task, showcasing the use of the 3D widget for translation, scaling, and rotation. 5 shows the final position and orientation after successfully completing the task.

We incorporated 3D widgets [103] for rotating and scaling tasks of the 3D object, following established practices and a conventional method for manipulation of 3D objects [144]. Our implementation of the 3D widgets can be seen in Figure 5.12 2-4.

The 3D cursor was designed to automatically pre-select the nearest 3D object or widget as participants moved it, as proposed by Baloup et al. [20]. Participants were then required to confirm the selection of the pre-selected object or widget, via touch. Upon selection, the 3D object had to be translated, with the translation being directly mapped to the cursor. Additionally, users interacted with 3D widgets for the necessary rotation and scaling of the object. We used axes separation [53, 163] on the 3D widgets, by which the cursor's up/down movements were scaling the 3D object uni-formally. Left/right movements were used to rotate the object around the y-axis. As in the first study, the cursor changed from a spherical shape (1 cm diameter) in the primary mode to a 3D crosshair (3 × 3 × 3 cm) in the refinement mode. For each trial, the position of the 3D target was randomly assigned from a pre-defined list of all possible positions. The list of target positions was generated before the study and included 3D coordinates which all had a different 3D position on all three axes. The rotation and scale of the 3D targets were randomized for each trial in a way that the next target was always at least $90 - 270°$ differently rotated and at least 10-30 cm differently scaled than the previous target. A docking task was deemed successful when the two objects (moving and target object) position aligned within $< 2cm$ on each axis, its rotation varied by $< 4°$ degrees, and its scale differed by $< 2cm$. We defined these parameters based on a previous pilot study and would like to highlight that this requires very high precision - when fulfilling these constraints, the two objects were visually perfectly aligned, as seen in Figure 5.12 5. As in the first study, we instructed participants to align the two 3D objects as precisely as possible and to do so as quickly as possible. To discourage excessive effort on accuracy at the expense of time, a limit of 40 seconds was placed on each trial.

Design

We employed a within-subject design involving repeated measures, focusing on two variables:

- primary technique *(Pointer | Hand | Head)*
- refinement technique *(None | Pointer | Hand | Head)*

Due to the incompatibility of the two techniques, the combination *Pointer+Head* was excluded. For each technique, participants completed a block of 3 docking tasks for each technique, in addition to 3 training blocks. In total, each participant completed 33 trials (11 techniques × 3 blocks). The order of interaction techniques presented to each participant was randomized.

We measured the task completion time of each trial. After each technique, participants were asked to provide their subjective feedback by commenting on the experienced primary and refinement technique regarding ease-of-use, physical, and mental demand. After the experiment, we asked participants to select and comment on the three best and worst-performing combinations. The subjective feedback phase was also used as a break from the interaction to mitigate potential physical fatigue.

Results

A repeated-measures ANOVA, with an alpha level set at .05, was conducted to analyze the task completion time data. Similar to the first study, we checked our data for sphericity and conducted post-hoc tests as part of our analysis. The time analyses included only successful target selections, accounting for 588 of 594 total trials, which represents 98.9% of the total data. Similarly to the first study, we considered all combinations where a particular technique occurred as a primary mode or refinement mode, respectively. It should be noted that all pairwise comparisons detailed below are significant at the $p < 0.001$ level, unless specified otherwise.

Time We found a significant main effect on placement time between the primary techniques ($F_{2,100} = 26.626, p < 0.001$), refinement techniques ($F_{3,150} = 7.535, p < 0.001$), the interaction between the primary × refinement techniques ($F_{6,300} = 2.339, p < 0.032$).

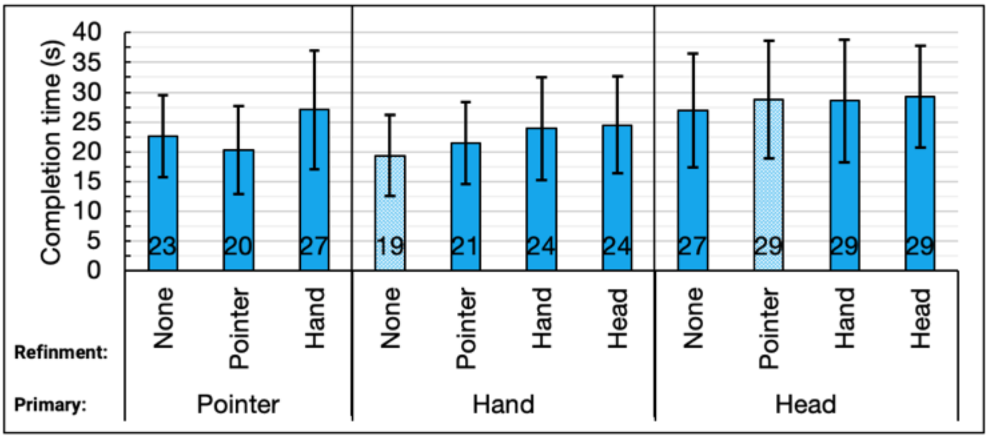


Figure 5.13: Mean placement task duration across our interaction techniques. Error bars indicating the standard deviations and the pattern-filled bars highlight the quickest and slowest techniques respectively.

In the comparison of primary techniques, it was found that both *Pointer* and *Hand* performed at a similar speed, and were faster than the *Head* condition, as shown in Figures 5.6 and 5.13. For the refinement techniques, we found that *Pointer* and *None*, were both equally fast and faster than *Hand* and *Head*, which were also equal to each other.

Looking at the interplay between primary × refinement techniques as depicted in Figure 5.8 (• dots), it is evident that the *Hand* technique stands out for its speed. Not only does it offer the fastest primary technique for rapid selection and translation of the 3D object, but it also maintains sufficient accuracy to be the overall quickest technique even without any refinement, as shown in the results of the *Hand-None* technique in Figure 5.8. Generally, we were surprised about the performance of Simo and our techniques, even when used without any refinement (*-*None* techniques). Despite the task requiring very precise 3D alignment of an object, participants accomplished it with ease. These findings lead us to conclude that the primary mode alone provided users with sufficient precision for fast 3D object alignment. However, we also need to take into account that the incorporation of mechanisms as 3D widgets, object pre-selection, and separation of axes for manipulation, helped a lot to achieve such positive results.

Subjective feedback Participants subjectively shared favorable feedback for *Hand-None* and *Hand-Hand*, expressed by comments like:

> *"Hand-None, It's very accurate, even without refinement, I would like that the refinement would be optional - triggered only once needed." - P6*

Pointer-Hand was also preferred by a few participants, with arguments as:

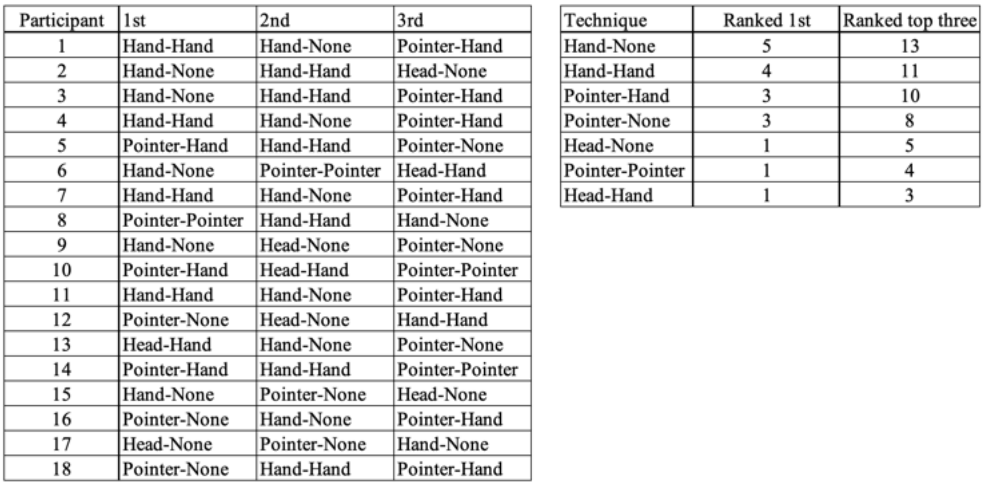
> *"Pointer-Hand, I like this, the other way around is however very bad." - P10*

Participants further supported the idea of refinement:

> *"Head-Pointer, selection, and translation are very fast, however, you need refinement for such small objects." - P14*

Not well-performing techniques participants argued with:

> *"Hand-Pointer, this is not good, I would rather use Hand-Hand." - P2*



Participant	1st	2nd	3rd
1	Hand-Hand	Hand-None	Pointer-Hand
2	Hand-None	Hand-Hand	Head-None
3	Hand-None	Hand-Hand	Pointer-Hand
4	Hand-Hand	Hand-None	Pointer-Hand
5	Pointer-Hand	Hand-Hand	Pointer-None
6	Hand-None	Pointer-Pointer	Head-Hand
7	Hand-Hand	Hand-None	Pointer-Hand
8	Pointer-Pointer	Hand-Hand	Hand-None
9	Hand-None	Head-None	Pointer-None
10	Pointer-Hand	Head-Hand	Pointer-Pointer
11	Hand-Hand	Hand-None	Pointer-Hand
12	Pointer-None	Head-None	Hand-Hand
13	Head-Hand	Hand-None	Pointer-None
14	Pointer-Hand	Hand-Hand	Pointer-Pointer
15	Hand-None	Pointer-None	Head-None
16	Pointer-None	Hand-None	Pointer-Hand
17	Head-None	Pointer-None	Hand-None
18	Pointer-None	Hand-Hand	Pointer-Hand

Technique	Ranked 1st	Ranked top three
Hand-None	5	13
Hand-Hand	4	11
Pointer-Hand	3	10
Pointer-None	3	8
Head-None	1	5
Pointer-Pointer	1	4
Head-Hand	1	3

Figure 5.14: The results of the subjective user rating. The left table shows individual participant ratings for interaction techniques, ranked from 1st to 3rd. The right table aggregates these ratings, indicating how often each technique was ranked 1st the number of times it appeared in the top three choices across all participants.

In terms of subjective ranting, *Hand-None* was ranked amongst the top three techniques by 13 participants and was ranked first 5 times. *Hand-Hand* was ranked 11 times in the top three, 4 times as first. *Pointer-Hand* was in the top three 10 times (3 times first). *Pointer-None* was 8 times in the top three, 3 times as first. *Head-None*, *Head-Hand*, and *Pointer-Pointer* were all ranked as first once and were in the top three 5, 3, and 4 times respectively. In Figure 5.14, we show the subjective user rankings across all participants and all interaction techniques.

5.3 Applications

In this section, we showcase various application scenarios demonstrating the advantages of using Simo for interacting with 3D content on a 2D display. Along the object manipulation methods discussed and evaluated in the study, we also explored other means of interactions that become possible through the available tracking and interaction possibilities.

3D Studio

The 3D Studio is an application designed to showcase the capabilities of the interaction techniques for pointing, selection, and manipulation of 3D content, as seen in the second study in a more crowded setting. As shown in Figure 5.15, users can freely rearrange and alter the interior of a room using it as a playground to test the different interaction techniques we presented in this work.

Besides this, we were also interested in how our system could be used to actively enrich the user-perception of a displayed 3D scene [282]. Due to the lack of depth-cues 3D objects or 3D scenes can just poorly be displayed on flat 2D screens. User-perspective rendering addresses this issue but usually requires complex hardware setups making such 3D experiences inaccessible for the mainstream [146]. Due to its head-tracking capabilities, Simo enables user-perspective rendering without any additional hardware and thus can significantly improve the 3D interaction and experience, as shown in Figure 5.16.

Free World Game

We were further interested if the body and head motion we capture could be used for navigation in a virtual world. In an outdoor game scene, we map the user's movement in physical space to the virtual avatar, see Figure 5.17. This enables the

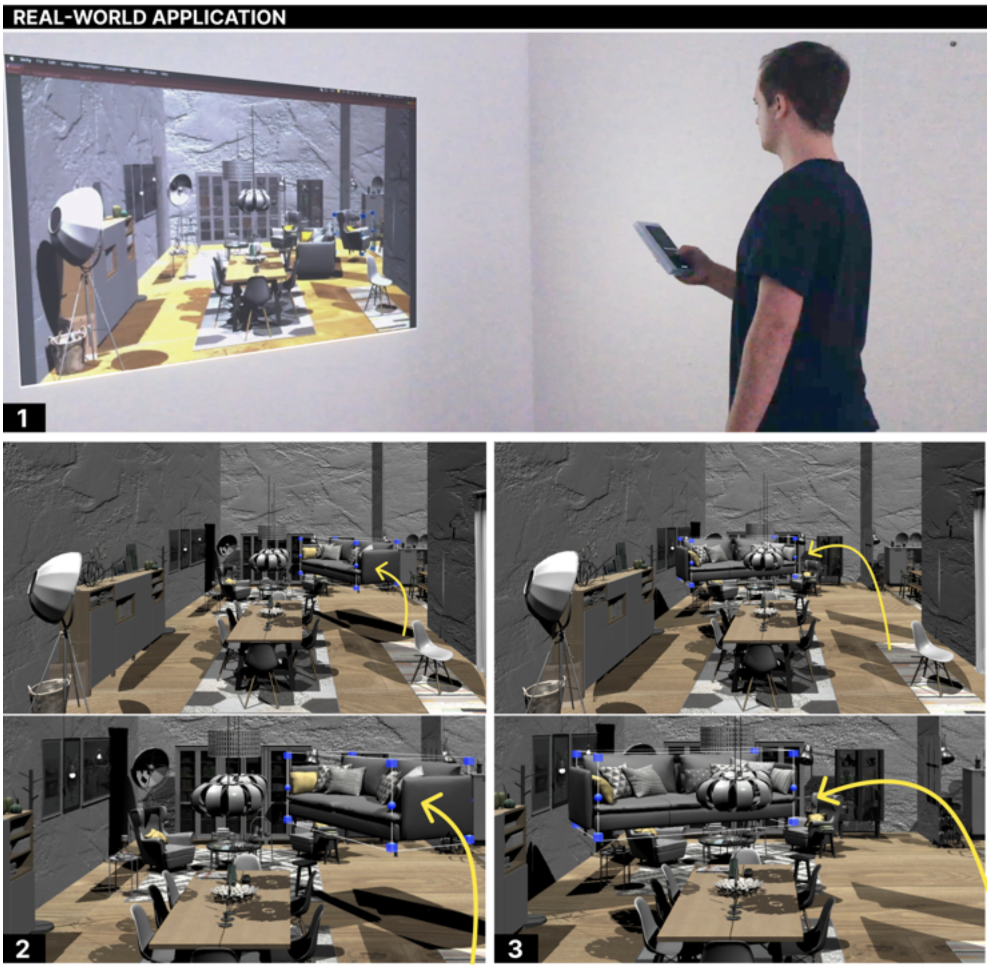


Figure 5.15: The user interacts with the 3D Studio app 1, where they can, for example, select a couch. Utilizing 3D manipulation 2, they then translate and rotate the couch, repositioning it to a new location in the virtual environment 3.

user to travel through the scene [281] and adapt their view, which, in combination with the user-perspective rendering from before, provides an even greater sense of immersion, even if the movement is restricted by the constraints of the physical room in the current implementation.

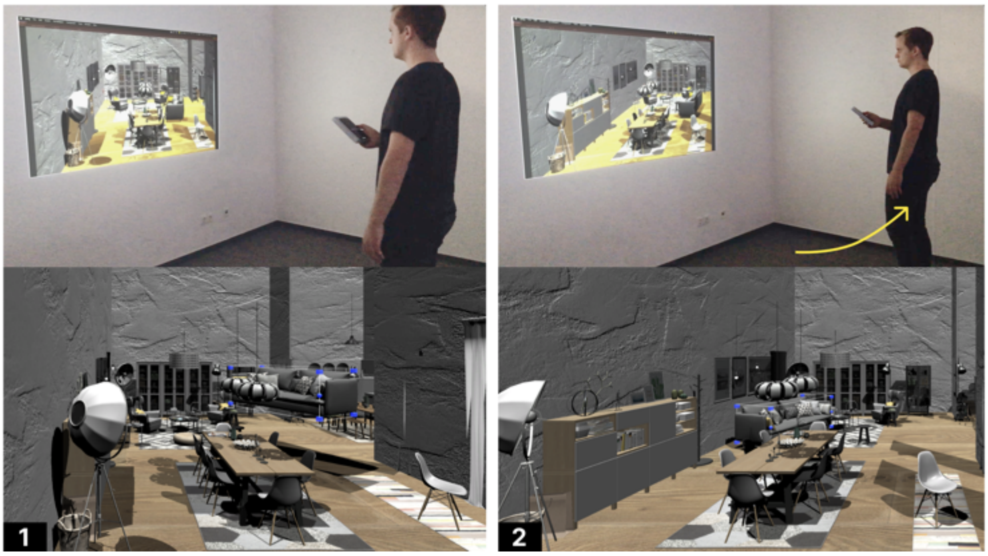


Figure 5.16: The user can move with its body around to adjust the view of the application, getting a better view of the 3D application. The initial position is shown in 1, while 2 represents an adjusted view, providing a better view into the 3D application from a different angle.

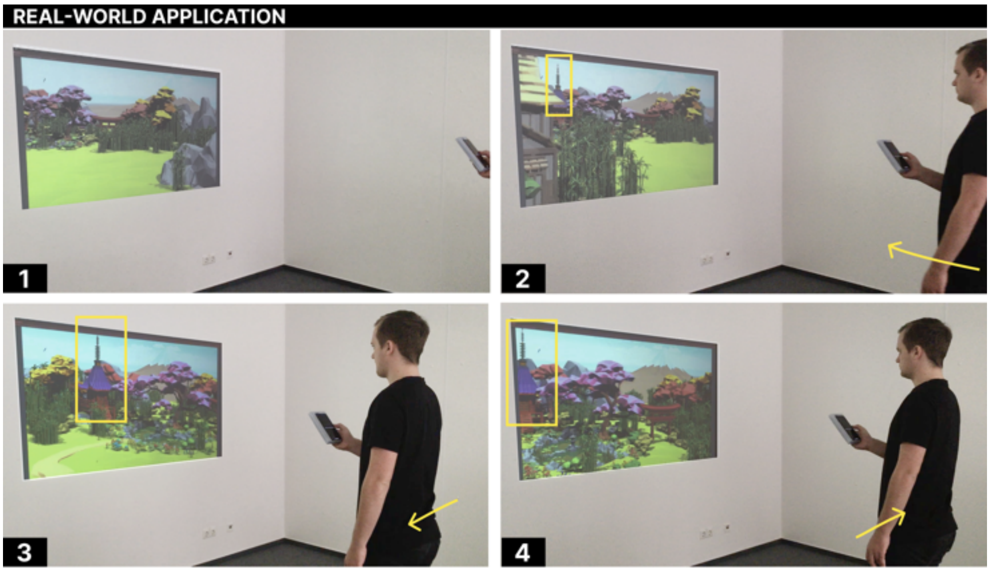


Figure 5.17: In this application user can move with its body around to adjust the view into the game application, however also the translation of the game camera or character is being adjusted by the user's position in front of the distant display. The image sequence shows: 1-2 the user doing a step forward, 3 the user does a step to the left (note how the position of the high building in the game annotated in the yellow square changes, due to the adapted game character position, which is mapped to the user's real-world position) and 4 the user does a step to the right.

5.4 Summary

In summary, our smartphone-based tracking approach enabled interaction techniques to achieve very satisfactory performance for 2D as well as 3D interactions, even without refinement and the need for complex tracking hardware. Many participants frequently emphasized this point, explaining that the fundamental primary techniques were often sufficiently accurate in most cases.

The findings from our studies indicate that by using a multimodal refinement mode, we can improve the pointing accuracy up to three times, beyond the accuracy of the primary techniques, without compromising time. It was observed that certain techniques are more effective as primary methods (for instance, *Head* and *Pointer*), while others excel in the refinement role (such as *Touch*). An in-depth comparison showed that by carefully combining different multimodal techniques, we can create high-performing techniques that are faster and, at the same time, also more accurate than any other combination within the same modalities, like the *Head-Touch* pairing for 2D interactions.

Our evaluation reveals that *Head-Touch* emerged as the best 2D interaction technique, excelling in both speed and accuracy, with *Head* and *Pointer* as the leading primary techniques, and *Touch* as the most effective refinement method. In 3D interaction, the *Hand-None* technique proved to be the fastest, while *Hand* and *Pointer* stood out as the most fastest primary techniques, and *None* and *Pointer* were the fastest refinement techniques.

Our findings also reveal that employing the same modality for both primary interaction and refinement proves effective, as shown in the results of combinations like *Pointer-Pointer* and *Hand-Hand*. Firstly, this approach is also viable for uni-modal systems, and secondly, our participants highlighted that switching between different motion types (such as *Pointer-Hand, Head-Pointer,* or *Hand-Pointer*) can lead to increased mental efforts and slower performance. Consequently, in certain scenarios, opting for a uni-modal combination may be a better choice.

In the following chapter, we present the discussion of the thesis.

6

Discussion

In this chapter, we present the discussion of the thesis. We begin with reflecting on the design challenges, objectives, and the main research question that guided this research. As we dive into the results of these challenges, we outline how this thesis and its project contributed to the challenges and also show how all design challenges collectively contributed to the overarching theme of reducing complexity and workload in spatial interactions. Finally, we offer critical insights and reflections on the subject matter of the thesis in the form of additional remarks.

6.1 Reflection of the Design Challenges

In the pursuit of creating more accessible spatial interaction systems for distant displays, this thesis addressed three design challenges. These challenges were not defined and later approached as mere obstacles, that need to be overcome, they have been seen as opportunities and inspirations to redefine how individuals engage with digital content and information presented on distant displays. By reimagining the systems we use to interact with distant displays, we aimed to minimize system complexity and maximize accessibility — the use for everyone.

6.1.1 Design Challenge: Tracking Device

Design Challenge 1: Tracking Device Options

The challenge was to determine whether the acquisition of dedicated devices is the sole viable option for users to effectively engage in and utilize spatial interactions with distant displays.

Contextualizing the Challenge

The first design challenge presented in this thesis revolved around exploring alternatives to dedicated devices for engaging in spatial interactions with distant displays. The fundamental question was whether users were confined to using specialized hardware, or if there were effective and accessible alternatives that could enable similar interactions.

Key Contributions to the Design Challenge

The development and implementations of Pocket6 and Simo, the smartphone-based spatial tracking systems, served as a direct response to this challenge. By leveraging the existing capabilities of regular smartphones, these implementations stood as evidence of the feasibility of using non-dedicated, personal, and ubiquitous devices for complex spatial interactions with distant displays.

- **Dual Camera Utilization** Simo's approach simultaneously used a smartphone's front and back cameras to enable comprehensive spatial tracking of the user's hand, head, body, and eye-gaze movements. This method provided a rich interaction framework without the need for additional, dedicated tracking hardware.
- **Versatility in Interaction Techniques** The system supported various interaction techniques such as touch, ray-casting, virtual hands, and head-pointing, demonstrated the versatility and adaptability of non-dedicated devices, placed in the environment or attached to the user, in facilitating diverse spatial interactions.
- **Technical Validation** Through the technical analyses and evaluations, Pocket6 and Simo validated their effectiveness in tracking head rotations, provided an understanding of the smartphone's front camera field of view, and determined the optimal device holding positions, underscoring the via-

bility of using smartphones as capable alternatives for specialized tracking devices.

- **Empirical Evaluation through User Studies** The user studies focusing on 2D and 3D interactions provided empirical evidence of user performance while using a common smartphone for complex spatial interactions. These studies directly compared and validated that non-dedicated devices could indeed match the interaction quality typically associated with specialized hardware.
- **Real-world Application Scenarios** The exploration of various application scenarios, such as the 3D design studios, gaming environments, and immersive navigation, further illustrated the potential of common smartphones to effectively handle tasks traditionally known for dedicated devices.

In conclusion, we believe that this thesis successfully addressed the first design challenge by demonstrating that the acquisition of dedicated devices was not the sole viable option for effective spatial interactions with distant displays. The development of Pocket6 and Simo underscored the potential of everyday smartphones to serve as powerful tools for spatial interaction, thereby widening accessibility and reducing the dependence on dedicated hardware. Our projects and explorations created new opportunities for more inclusive and readily available spatial interaction technologies, democratizing access to advanced human-computer interaction paradigms.

6.1.2 Design Challenge: Impact of Hardware Removal

Design Challenge 2: Impact of Hardware Removal

The challenge was to determine whether the elimination of dedicated hardware devices from spatial interaction systems unintentionally leads to a reduction in the effectiveness of established interaction techniques or impairs the overall user interaction experience.

Contextualizing the Challenge

The second design challenge questioned whether eliminating dedicated hardware in favor of non-specialized devices, like smartphones, would compromise the effectiveness of spatial interaction techniques and overall user experience. This challenge was rooted in the concern that transitioning away from specialized equipment might lead to diminished interaction options or user interaction performance.

Key Contributions to the Design Challenge

- **Preservation of Interaction Techniques** Pocket6 and Simo demonstrated that established interaction techniques like touch, ray-casting, virtual hand, and head-pointing could be effectively adapted to non-dedicated devices. The thesis showed these techniques were not only preserved but were also effectively executed using a smartphone, maintaining their core functionalities and benefits.
- **User Study Outcomes** The user studies conducted to evaluate 2D and 3D interactions provided imperative insights. They indicated that users could achieve high precision and fast interactions with Pocket6 and Simo, comparable to those expected from specialized hardware. These studies showed that the user experience and interaction efficiency were not only preserved but, in some cases, also enhanced.
- **Feedback and User Acceptance** Participant feedback from the studies illustrated a high level of acceptance and satisfaction with the smartphone-based interaction system. Users appreciated the seamless integration of familiar touch interactions with more novel spatial techniques, highlighting an overall positive user experience.

In addressing the second design challenge, this thesis provided considerable evidence that moving away from dedicated hardware devices to more ubiquitous and accessible technology like smartphones did not inherently reduce the effectiveness of spatial interaction techniques or degrade the user experience. On the contrary, the projects in the thesis exemplified how users could retain, and in some aspects enhance, the quality of their interaction with distant displays by leveraging the widespread familiarity and advanced capabilities of modern smartphones. This smartphone approach not only maintained the inclusion of established interaction paradigms but also democratized access to sophisticated interaction techniques, potentially leading to broader adoption and innovation in the field of human-computer interaction.

6.1.3 Design Challenge: Personal Device Feasibility

Design Challenge 3: Personal Device Feasibility

The challenge was to assess the feasibility of using a personal device to enable spatial interactions and determine the extent of its spatial interactive capabilities.

Contextualizing the Challenge

The third design challenge focused on evaluating the practicality of utilizing personal devices, specifically smartphones, for spatial interactions. It questioned the extent to which everyday devices such as smartphones could facilitate spatial interactions, examining their capabilities and limitations in various contexts.

Key Contributions to the Design Challenge

- **Continuous Evaluation Across all Projects** Throughout our research project, we continuously focused on user evaluation across all our thesis projects, where the primary objective was to explore the exact extent of input modalities that smartphones can support. This research involved an investigation of how many input modalities could be enabled and what interaction techniques could be created. Furthermore, it encompassed an evaluation of all these techniques in various studies and 2D and 3D tasks, with attention on metrics such as error rates, response times, and subjective user feedback. This comprehensive evaluation provided detailed insights into the exact extent of interactive possibilities enabled by smartphone-based spatial tracking.
- **Refinement of Techniques** Additional methodologies, beyond input modalities and interaction techniques, such as the implementation and evaluation of refining modes, played a crucial role in outlining the extent of smartphone-based interaction possibilities.

In conclusion, we believe that this thesis successfully addressed the third design challenge by the extensive exploration into the feasibility of smartphones for spatial interactions. The emphasis on evaluating input modalities, developing innovative interaction techniques, and conducting extensive studies allowed for a comprehensive understanding of the extent of interaction possibilities.

6.2 Reflection of the Thesis Objectives

In this section, we reflect on the objectives that guided our research throughout the thesis, providing an overview of how each objective was addressed and the key contributions made in pursuit of these goals.

6.2.1 Objective 1: Analyzing Spatial Interaction Systems

Identifying drawbacks and challenges as well as opportunities of existing spatial interaction systems.

Interactive systems enabling spatial interactions are not a novel device category and have been around for almost 30 years. Hence, to improve the usability and usefulness of this device category, the first objective was to analyze the status quo. This analysis informed all the next steps of this work.

Context of the Objective

Objective 1 centered on a comprehensive analysis of the landscape of spatial interaction systems. Its aim was to identify the drawbacks, challenges, as well as the opportunities inherent in these systems. This objective served as a foundation for informing future developments and improvements in the field.

Approach and Implementation in the Thesis

- **Review of Existing Systems** The thesis began with the creation of a detailed taxonomy of existing spatial interaction systems. This review encompassed a range of systems, including dedicated devices and setups used for spatial interactions in various contexts. The examination of these systems led to the identification of common challenges, such as the need for complex hardware, limited accessibility, and constraints in interaction techniques.
- **Identification of Challenges and Drawbacks** Key challenges were identified, including reliance on expensive and cumbersome hardware, a steep learning curve for users, and limitations in the versatility and adaptability of the systems. These drawbacks were crucial in understanding the barriers to the widespread adoption and effectiveness of spatial interaction systems.
- **Exploration of Opportunities** Alongside challenges, the thesis also ex-

plored opportunities presented by emerging technologies and changing user needs. For instance, the increasing ubiquity of smartphones and advancements in their capabilities presented a significant opportunity to leverage these personal devices for spatial interactions.

- **Comparative Analysis** A comparative analysis was conducted between traditional spatial interaction systems and the proposed smartphone-based system. This analysis highlighted areas where conventional systems fell short and where smartphone-based systems could offer improvements in accessibility, usability, and cost-effectiveness.
- **Feedback and Insights from User Studies** User studies conducted as part of the thesis provided real-world insights into the user experience with spatial interaction systems. Feedback from these studies proved valuable in understanding user expectations, preferences, and the practical challenges faced while interacting with these systems.

In addressing the first objective, the thesis provided a thorough analysis of the state of spatial interaction systems, identifying key challenges and opportunities. This analysis served as a critical foundation for subsequent developments and improvements in these systems, especially in the context of using everyday personal devices like smartphones. The insights gained from this objective informed the direction of future research and development in spatial interactions, aiming to make them more accessible and versatile.

6.2.2 Objective 2: Developing a Smartphone-Based Solution for Spatial Interactions

Investigating strategies to overcome the identified issues and further improve existing advantages by developing a new smartphone-based software solution for spatial interactions with distant displays.

Based on the issues and opportunities identified in Objective 1, we planned to investigate potential solutions. This investigation included (1) the design of multimodal interaction techniques, (2) an implementation that comprehensively addressed the problem, (3) a thorough evaluation of the developed solution, and (4) example applications demonstrating the system's usage in various scenarios. The objective aimed to discover new ways to enable handheld-based spatial interaction without compromising tracking capability or application usability.

Context of the Objective

Objective 2 focused on inventing strategies to address the challenges identified in Objective 1, while simultaneously enhancing the existing benefits of spatial interaction systems. The primary aim of this objective was the development of a new smartphone-based software solution, specifically tailored for spatial interactions with distant displays.

Approach and Implementation in the Thesis

1. **Design of Multimodal Interaction Techniques** The thesis placed significant emphasis on implementing various interaction techniques that are adaptable to a smartphone-based system. These techniques were unimodal as well as multimodal. They leveraged different input modalities, such as touch, hand gesture, and head movement, to offer a comprehensive spatial interaction. The design process involved identifying how these modes can be intuitively integrated into a smartphone interface.
2. **Comprehensive Implementation of the Solution** The implementation phase involved creating a software solution that effectively utilizes the capabilities of modern smartphones for spatial interaction. This included leveraging built-in sensors and cameras for motion tracking and a touchscreen for direct touch inputs and manipulation. The software was developed to be robust, open-source, reusable, and capable of handling various interaction paradigms.
3. **Thorough Evaluation of the Solution** An extensive evaluation of the developed solution was conducted to assess its effectiveness and usability. This involved user studies where participants interacted with distant displays using the newly developed system. The evaluation focused on aspects like speed, precision of interaction, user satisfaction, and overall performance compared to traditional systems.
4. **Demonstration through Example Applications** The thesis showcased the practical utility of the system through various application scenarios. These demonstrations highlighted how the system could be used in different contexts, ranging from simple 2D menu selection tasks to complex 3D interactions. The examples served as proof of concept for the versatility and

adaptability of the smartphone-based solution.

Addressing Objective 2, the thesis successfully developed and shared a novel smartphone-based solution for spatial interactions. This solution overcame many issues identified with traditional systems and enhanced their advantages using the ubiquity and capabilities of smartphones. Through detailed design, implementation, evaluation, and demonstration, the thesis established a robust framework for future advancements in spatial interaction technology. It emphasized handheld-based interactions without compromising tracking capabilities or application usability, marking a significant step in making spatial interactions more accessible and adaptable for a wider user base and applications.

6.2.3 Objective 3: Generalizing Findings for Spatial Interaction Systems Designers and Developers

Generalizing the collected findings in ways so that they can be helpful for designers and developers, who are creating applications and systems for spatial interaction.

The main goal of this objective was to find ways to present and communicate the findings from the previous objectives, which were not tightly connected to a particular use case or scenario. The aim was to describe the impact of specific design decisions in a general manner.

6.2.4 Context of the Objective

Objective 3 of the thesis was to extrapolate and generalize the findings from previous objectives. The aim was to articulate these findings in ways that are applicable and valuable for designers and developers in spatial interaction applications and systems, regardless of specific use cases or scenarios.

6.2.5 Approach and Implementation in the Thesis

- **Synthesizing Key Insights** The thesis involved synthesizing the key insights and learnings from the development, implementation, and evaluation phases.
- **Highlighting the Impact of Design Decisions** A crucial part of the thesis was to clearly articulate the impact of specific design decisions. This included

an exploration of how different interaction modalities, user interface designs, and tracking technologies affect user experience, system performance, and application flexibility.

- **Communicating Transferable Concepts** The thesis emphasized on communicating concepts and findings in a way that they can be transferred to other applications and systems. This involved identifying and describing the underlying principles that influence effective spatial interaction, regardless of the particular technology or application domain.
- **Providing a Future Work Directions** Based on our experience and findings, we proposed future work directions. This served as a reference point for designers and developers, helping them make informed decisions and innovate within their specific contexts.

Objective 3 was successfully achieved through a comprehensive approach that extracted, generalized, and communicated the findings from the research in a universally applicable manner. The thesis provided a valuable resource for designers and developers in the field of spatial interaction, offering guidance grounded in thorough research and practical experimentation. The presented findings equipped practitioners with the knowledge and tools necessary to innovate and excel in the domain of spatial interaction.

6.3 Reflecting the Main Research Question of the Thesis

In the following, we reflect on the main research question that has guided this thesis, delving deeper into its fundamental aspects and implications. We critically examine the main research question, reflecting on how the insights and developments presented contribute to its understanding and resolution."

Main Research Question of this Thesis

Can we build a spatial interaction system for distant displays that everyone access and use without requiring a dedicated hardware device?

6.3.1 Context of the Main Research Question

The central research question of the thesis asked whether it was possible to develop a spatial interaction system for distant displays that is universally accessible, eliminating the need for specialized hardware devices. This inquiry was pivotal in democratizing spatial interaction technologies, making them available to a broader audience without the barriers posed by dedicated hardware.

6.3.2 Approach and Findings in the Thesis

- **Accessibility and Universal Design Considerations** The system was designed with a strong focus on the universal design principle. This approach ensured that the system was not only technically feasible but also adaptable to the diverse needs and preferences of a wide range of users and contexts.
- **Development of Smartphone-Based Solution** The core of the thesis revolved around leveraging the ubiquity and advanced capabilities of modern smartphones to develop a spatial interaction system. The project capitalized on the existing sensors and cameras in modern smartphones to enable precise tracking and interaction capabilities, typically reliant on specialized hardware.
- **Innovative Tracking and Interaction Techniques** The thesis introduced and implemented innovative tracking techniques, such as the simultaneous use of front and back cameras for comprehensive spatial sensing. This was complemented by the development of multimodal interaction techniques, integrating touch inputs with real-time tracking of hand, head, and eye movements.
- **Technical Feasibility and User-Centric Evaluation** Extensive technical evaluations were conducted to ascertain the feasibility and effectiveness of the smartphone-based system. These evaluations encompassed various interaction paradigms, from simple 2D pointing tasks to complex 3D object manipulations, ensuring the system's applicability across diverse use cases.
- **Empirical Validation through User Studies** The research included comprehensive user studies to validate the system's effectiveness and user satisfaction. These studies provided empirical evidence that the smartphone-based system can indeed facilitate efficient and effective spatial interactions

for distant displays without requiring dedicated hardware.

The research successfully answered the main question by demonstrating that it is indeed feasible to build a spatial interaction system for distant displays accessible to everyone, without the need for dedicated hardware devices. The thesis achieved this through the innovative use of commonly available smartphones, leveraging their built-in sensors and capabilities to create a versatile and effective spatial interaction platform. This achievement marked a significant step forward in making spatial interaction technologies more inclusive, accessible, and adaptable to various contexts and user needs.

6.4 Further Discussion

In this following section, we step back to share personal reflections, opinions, and the broader implications stemming from the research journey undertaken in this thesis. It is a space where subjective insights and standpoints are brought forward, adding depth and personal context to the scientific discourse.

6.4.1 Contemporary Relevance and Impact of the Developed System

In the implementation chapters of our project, we progressively improved our smartphone-based system. Initially, we envisioned a single-phone tracking system, but technical limitations in iOS and ARKit at the time made this unfeasible. Consequently, we adopted a two-phone prototype approach, focusing on designing interaction principles and evaluations primarily enabled by smartphone camera-based spatial tracking. As iOS and ARKit evolved, they eventually supported the simultaneous use of front and back cameras on a single device, which was essential for our tracking needs. This development was a significant step forward, aligning closely with our original single-phone concept. By this, our vision became much more realistic since this opened up the possibility that anyone could simply download a Simo app and use their smartphone for advanced spatial interactions. Even though this advancement eliminated the need for a second smartphone device for camera functions, we still view our novel implementations, algorithms, evaluations, and applications as unique and central contributions of our work. Additionally, it is important to note that creating Simo or a similar system for Android, which dominates over 70% of the global market, presents many challenges and limitations. As Android's ARCore currently lacks support for simultaneous tracking, researchers working with Android may need to rely on a two-phone setup, explained in detail in this work, for exploratory and research activities.

6.4.2 Spatial Interaction Devices within the HCI Research Community

In our research journey, we have noticed a lack of focus on universal design within the Human-Computer Interaction (HCI) community. This issue stems from unclear and often conflicting definitions of terms like accessibility and diversity, as we discussed in our introductory chapters. HCI, at its core, is about embracing user diversity and including everyone. Yet, surprisingly, there seems to be minimal research dedicated to defining and applying universal design in HCI contexts. This trend is also evident in research projects focusing on spatial interaction, devices, and interaction techniques. Our analysis shows that most of these projects aim to develop new devices and techniques, often seeking to create something completely new and unique. These projects typically last several months, aligning with the deadlines for biannual conference submissions. Over the last two decades, this pattern has not changed much. For example, we picked out two projects, Soap from Baudisch et al. [24] and SurfAirs from Courtoux et al. [56] depicted in Figure 6.1, 17 years apart, both aimed at improving interactions with distant displays. However, both focused more on practical, even do-it-yourself, device innovation. While these projects are undoubtedly valuable, they seem to represent a dominant trend in the field. We believe that longer projects, those spanning over multiple months or even years, might not appeal to many researchers due to their lengthy duration and conflict with frequent conference submission deadlines. As a result, these projects often do not advance beyond basic implementations. It is rare for projects exploring new input devices and interaction techniques to deeply refine and perfect their concepts. For instance, once a basic interaction like ray-casting with a handheld device is working, the project often moves to the evaluation phase instead of improving finer details like signal filtering, incorporation of UI widgets, cursor object pre-selection mechanisms, or adding refinement modes. These deeper dives into development, such as making the cursor movement smoother, might take weeks or months, with the outcome "just" of a nicer cursor movement. We believe, however, that this is essential and can significantly enhance the interaction quality. Based on our experience, we can confidently say that putting in this extra effort can greatly improve the system's performance and the depth of its evaluation.

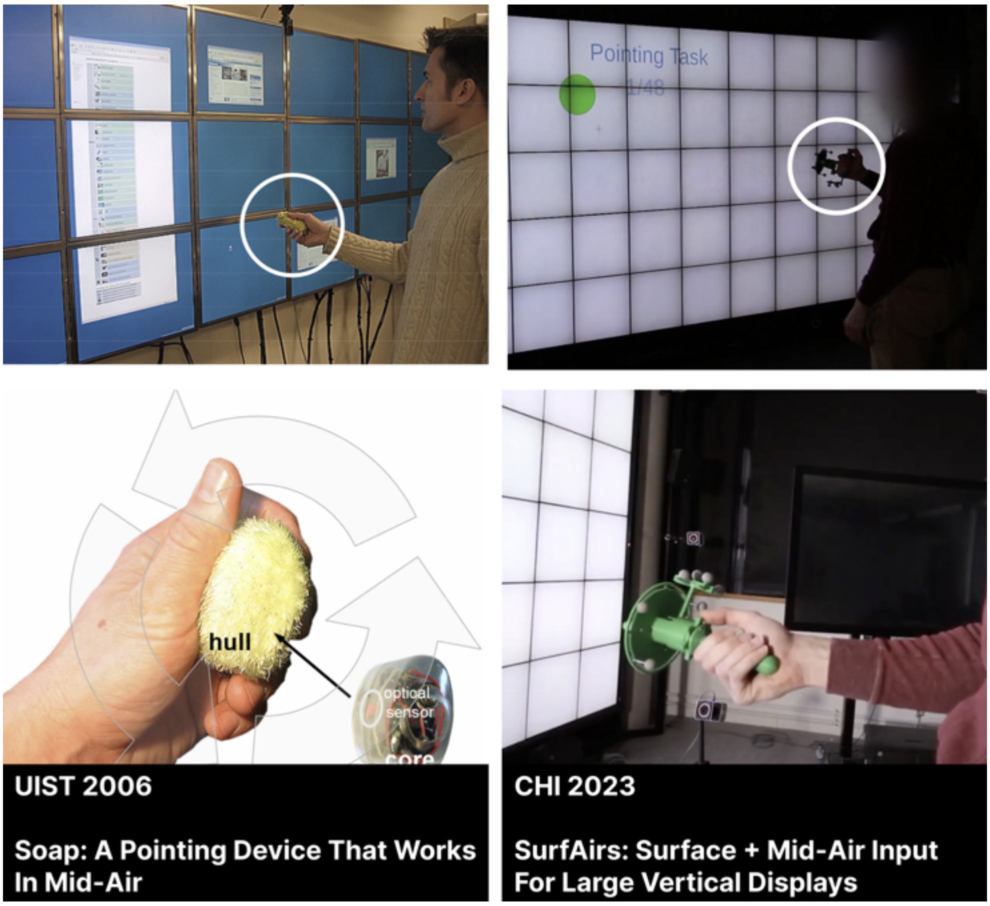


Figure 6.1: Example of two projects, although 17 years apart, adopted a similar strategy of developing new handheld hardware input devices for improving the interaction with distant displays.

6.4.3 Embracing the Cyborg Reality: The Transformative Role of Smartphones

Elon Musk's thought-provoking statement, "We are already cyborgs!" resonates deeply as we explore the realm of personal devices, especially smartphones, for innovative purposes. Our research journey started with a focus on specialized hardware but gradually shifted to these ubiquitous personal devices.

Reflecting on this shift, it becomes clear that, as HCI researchers, we might have underestimated the significance of smartphones. It is essential to recognize the vast scope of smartphone usage, with over 6.5 billion smartphones globally, reaching 85% of the world's population. This massive adaptation of an interactive device highlights the untapped potential in the HCI field, particularly in repurposing smartphones for diverse applications. Imagine conducting large-scale, real-world studies simply through a smartphone app download.

Moreover, our perspective on the relationship between humans and technology needs to evolve. Following Musk's insight, it is possible that we are edging closer to a cyborg-like existence, given our constant interaction with smartphones. They are with us, at us, at all times — while we work, rest, and even sleep. This continuous attachment offers a new lens through which to view the human-technology interface, one that is crucial for future HCI research.

7

Conclusion & Future Work

To conclude, when compared to existing research in the field, the findings of this dissertation offer both confirmations and novel insights. While some results align with previous research, the unique contributions of this dissertation — particularly the repurposing of the smartphone — provide new directions for the field.

This thesis represented a confluence of exploration and innovation in the realm of Human-Computer Interaction (HCI), particularly focusing on spatial interactions with distant displays. It ventured beyond traditional boundaries, challenging the status quo and navigating through the intricacies of integrating modern technology with user-centric design.

Our journey began with an acknowledgment of the existing limitations in spatial interaction devices and techniques. The path led us through the exploration of novel solutions, primarily revolving around the novel capabilities of modern smartphones. This exploration was not just about leveraging existing technology but about reimagining its potential in ways previously unconsidered.

The thesis underscored the ubiquitous nature of smartphones, highlighting their untapped potential in HCI. We delved into the development of a smartphone-based software solution for spatial interactions, embracing the idea that the most common personal device could be transformed into a powerful tool for interaction with distant displays.

Significant contributions of this work include the design of multimodal interaction techniques, the development of a comprehensive software solution, and thorough evaluations that tested the effectiveness of these innovations. The journey advanced in demonstrating the practical applications of our research, showcasing the versatility and adaptability of the developed solutions.

This thesis was more than a culmination of research; it was a beginning. It opened up avenues for further exploration in HCI, particularly in how we perceive and utilize personal devices like smartphones. It challenged future researchers to think beyond conventional uses and to envision a world where our daily gadgets become integral to advanced spatial interactions. The journey thus far is a testament to the potential that lies in rethinking the norm, a potential that is boundless and waiting to be unlocked.

In conclusion, this thesis is a reflection of a journey - one that started with questions and led to discoveries, innovations, and more questions. It was a journey that does not end with these pages but hopefully continues in the hands of future researchers, designers, and innovators who will take these ideas and forge new paths in the ever-evolving landscape of HCI.

7.1 Limitations

While Pocket6 and Simo demonstrated promising capabilities in spatial tracking and interaction, they were not without limitations. These limitations highlighted the challenges in the applications' robustness, ecosystem constraints, and environmental dependencies, setting a stage for future improvements and research directions.

- **Device Variability** A notable constraint of Pocket6 was the device variability. Systems like Pocket6 depended on the smartphone's inherent features and specifications, which varied across different models and brands. The initial implementation primarily utilized Apple's iPhone ecosystem, compatible with ARKit-supported devices. However, performance and experience might have varied across different smartphone models and operating systems, such as Android. Exploring a wider range of devices could have provided a more comprehensive understanding of the system's adaptability to various devices.
- **Environmental Conditions** The environmental factors such as lighting conditions and physical space also affected Pocket6's performance. The

experiments were conducted under controlled conditions with adequate lighting and without physical obstructions, such as objects placed around the user. Real-world scenarios might present varied lighting conditions and physical constraints, potentially impacting the performance and applicability of Pocket6.

- **Tracking Limitations** The tracking capabilities for head rotation in our systems, including the field of view and the smartphone holding position, were effective but had limitations in terms of the range of motion that could be reliably tracked. This impacted interactions that depended heavily on accurate tracking and potentially limited the system's utility in applications that required broader head movement tracking beyond our experiment's scope.
- **Study Design and Task Complexity** Finally, the design of our study tasks focused on specific interaction modalities, such as 2D pointing and 3D object manipulation for distant displays. Expanding even more advanced task complexities and varieties could provide deeper insights into the system's versatility and applicability across a broader range of applications and scenarios.

7.2 Future Work

The thesis research led to the development of a diversity of interactive systems, interaction techniques, and use cases. While this showed the rich possibilities of the proposed spatial interactions, there are still many concrete directions into which our principle could be improved or further investigated:

- **Adding feedback** The smartphone-based interaction, more concretely, the underlying smartphone application, could be enhanced to leverage smartphone-based haptic feedback. Supporting additional sensory feedback modalities such as haptic feedback was already proven beneficial, therefore this should be added to approaches as ours as well [203]. Besides haptic feedback also auditory feedback would be highly interesting to investigate, be it for tap/click events but also in the context of the moving hand as real-time spatial audio feedback, be it from the smartphone itself or spatial headphones (for example, Apple AirPod's spatial audio functionality).

- **Adding additional interaction modalities** Another interesting direction would be adding additional input modalities or additional axes for the input's degrees of freedom. These could be, for example, using the smartphone's touch pressure sensitivity as an additional axis when grabbing a virtual object. Also, an interesting direction would be the addition of speech input for voice commands [62, 168].
- **On-the-go interaction** While we explored our interactive system in a controlled setting, it could be can also explored in situations where users are on-the-move, while walking or walk-by interactions [127]. Since our systems already free the user from the usual constraints of fixed interaction spaces, it would be interesting to explore where are the limits in that regard [205]
- **Multi-user application scenarios** Another valuable direction would be the usage of smartphone-based tracking approaches in collaborative 3D object manipulation scenarios [80, 169], spatial design ideation scenarios [263] or cross-device interaction scenarios [189]. In current setups, where cameras need to observe users (where a user can always hide behind another user) or multiple cables need to be connected through the room, the smartphone-based approach could be beneficial. Even a very large collaborative model could be tested. Hereby, we mean really more than 50, 100, or even more users who download the spatial tracking app and collaborate. For example, in a Minecraft-like world.
- **Adding other smartphone functionalities** Beside the user motion tracking capabilities, which we used, also other smartphone AR features could be investigated and potentially added. This could be novel ways to generate and share the AR world-map (tracking features) among users by multi-user or online stored world-maps (cloud anchors), as well as other user tracking features like front and rear-camera skeleton tracking, real-world 3D even object scanning and interaction.
- **Smartphone + AR Glasses usage** Finally, another interesting future research direction would be to investigate highly fluid and device-independent spatial interaction principles and interactions. While in our work, we focus on distant display and smartphone spatial interaction scenarios, it would be interesting to find out how we can create a smooth transition between smartphone-only use (handheld AR), to distant display use, to AR glasses use-cases.

A

Appendix A: Supplementary Material

A.1 Questionnaires

Worst / am **schlechtesten** ☹

Best / am **besten** 😊

1	2	3	4	5	6	7

Easiness / Mühelosigkeit

IN-FRONT (davor)							
HANDS-DOWN (hände runter)							
SITTING (sitzend)							
CONTROLLER							

Fatigue / Erschöpfung

IN-FRONT (davor)							
HANDS-DOWN (hände runter)							
SITTING (sitzend)							
CONTROLLER							

Speed / Schnelligkeit

IN-FRONT (davor)							
HANDS-DOWN (hände runter)							
SITTING (sitzend)							
CONTROLLER							

Precision / Genauigkeit

IN-FRONT (davor)							
HANDS-DOWN (hände runter)							
SITTING (sitzend)							
CONTROLLER							

General impression / Gesamteindruck

IN-FRONT (davor)							
HANDS-DOWN (hände runter)							
SITTING (sitzend)							
CONTROLLER							

Figure A.1: Questionnaires used in the subjective evaluation of Pocket6 in Chapter 3.

List of Figures

Bibliography

[1] Sandip Agrawal, Ionut Constandache, Shravan Gaonkar, Romit Roy Choudhury, Kevin Caves, and Frank DeRuyter. “Using mobile phones to write in air”. In: *Proceedings of the 9th international conference on Mobile systems, applications, and services - MobiSys ’11*. the 9th international conference. Bethesda, Maryland, USA: ACM Press, 2011, p. 15.

[2] Roland Aigner, Daniel Wigdor, Hrvoje Benko, Michael Haller, David Lindlbauer, Alexandra Ion, Shengdong Zhao, and Jeffrey Koh. *Understanding Mid-Air Hand Gestures: A Study of Human Preferences in Usage of Gesture Types for HCI*. Nov. 2012.

[3] Florian Alt, Andreas Bulling, Gino Gravanis, and Daniel Buschek. “GravitySpot: Guiding Users in Front of Public Displays Using On-Screen Visual Cues”. In: *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology - UIST ’15*. the 28th Annual ACM Symposium. Daegu, Kyungpook, Republic of Korea: ACM Press, 2015, pp. 47–56.

[4] Florian Alt, Alireza Sahami Shirazi, Thomas Kubitza, and Albrecht Schmidt. “Interaction techniques for creating and exchanging content with public displays”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’13. Paris, France: Association for Computing Machinery, Apr. 27, 2013, pp. 1709–1718.

[5] Jessalyn Alvina, Carla F. Griggio, Xiaojun Bi, and Wendy E. Mackay. “CommandBoard: Creating a General-Purpose Command Gesture Input Space for Soft Keyboard”. In: *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology - UIST ’17*. the 30th Annual ACM Symposium. Quebec City, QC, Canada: ACM Press, 2017, pp. 17–28.

[6] Apple. *ARKit*. URL: https://developer.apple.com/augmented-reality/arkit/.

[7] Apple. *ARKit - World Tracking*. URL: https://developer.apple.com/documentation/arkit/understanding_world_tracking.

[8] Cagan Arslan, Yosra Rekik, and Laurent Grisoni. “E-Pad: Large Display Pointing in a Continuous Interaction Space around a Mobile Device”. In: *Proceedings of the 2019 on Designing Interactive Systems Conference*. DIS ’19. San Diego, CA, USA: Association for Computing Machinery, June 18, 2019, pp. 1101–1108.

[9] Clemens Arth, Raphael Grasset, Lukas Gruber, Tobias Langlotz, Alessandro Mulloni, Dieter Schmalstieg, and Daniel Wagner. “The History of Mobile Augmented Reality”. In: (2015), p. 41.

[10] ARToolKit. URL: http://www.hitl.washington.edu/artoolkit/.

[11] Alec Azad, Jaime Ruiz, Daniel Vogel, Mark Hancock, and Edward Lank. “Territoriality and behaviour on and around large vertical publicly-shared displays”. In: *Proceedings of the Designing Interactive Systems Conference on - DIS ’12*. the Designing Interactive Systems Conference. Newcastle Upon Tyne, United Kingdom: ACM Press, 2012, p. 468.

[12] Teo Babic, Harald Reiterer, and Michael Haller. “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, Oct. 13, 2018, pp. 2–10.

[13] Huidong Bai, Lei Gao, Jihad El-Sana, and Mark Billinghurst. “Free-hand interaction for handheld augmented reality using an RGB-depth camera”. In: *SIGGRAPH Asia 2013 Symposium on Mobile Graphics and Interactive Applications on - SA ’13*. SIGGRAPH Asia 2013 Symposium on Mobile Graphics and Interactive Applications. Hong Kong, Hong Kong: ACM Press, 2013, pp. 1–4.

[14] Gilles Bailly, Jörg Müller, Michael Rohs, Daniel Wigdor, and Sven Kratz. “ShoeSense: a new perspective on gestural interaction and wearable applications”. In: *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI ’12*. the 2012 ACM annual conference. Austin, Texas, USA: ACM Press, 2012, p. 1239.

[15] Matthias Baldauf, Florence Adegeye, Florian Alt, and Johannes Harms. “Your browser is the controller: advanced web-based smartphone remote controls for public screens”. In: *Proceedings of the 5th ACM International Symposium on Pervasive Displays - PerDis ’16*. the 5th ACM International Symposium. Oulu, Finland: ACM Press, 2016, pp. 175–181.

[16] Matthias Baldauf, Markus Salo, Stefan Suette, and Peter Fröhlich. “Display pointing: a qualitative study on a recent screen pairing technique for smartphones”. In: *Proceedings of the 27th International BCS Human Computer Interaction Conference*. BCS-HCI ’13. London, UK: BCS Learning & Development Ltd., Sept. 9, 2013, pp. 1–6.

[17] Robert Ball, Chris North, and Doug A. Bowman. “Move to improve: promoting physical navigation to increase user performance with large displays”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI ’07*. the SIGCHI Conference. San Jose, California, USA: ACM Press, 2007, pp. 191–200.

[18] Rafael Ballagas, Michael Rohs, and Jennifer G. Sheridan. “Sweep and point and shoot: phonecam-based interactions for large public displays”. In: *CHI ’05 Extended Abstracts on Human Factors in Computing Systems*. CHI EA ’05. Portland, OR, USA: Association for Computing Machinery, Apr. 2, 2005, pp. 1200–1203.

[19] Rafael Ballagas, Michael Rohs, Jennifer G. Sheridan, and Jan Borchers. “BYOD: Bring Your Own Device”. In: *In Proceedings of the Workshop on Ubiquitous Display Environments, Ubicomp*. 2004.

[20] Marc Baloup, Thomas Pietrzak, and Géry Casiez. “RayCursor: A 3D Pointing Facilitation Technique based on Raycasting”. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI ’19*. the 2019 CHI Conference. Glasgow, Scotland Uk: ACM Press, 2019, pp. 1–12.

[21] Domagoj Baricevic, Tobias Hollerer, Pradeep Sen, and Matthew Turk. “User-Perspective AR Magic Lens from Gradient-Based IBR and Semi-Dense Stereo”. In: *IEEE Transactions on Visualization and Computer Graphics* 23.7 (July 1, 2017), pp. 1838–1851.

[22] Domagoj Baričević, Tobias Höllerer, Pradeep Sen, and Matthew Turk. “User-perspective augmented reality magic lens from gradients”. In: *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology - VRST ’14*. the 20th ACM Symposium. Edinburgh, Scotland: ACM Press, 2014, pp. 87–96.

[23] Mayra Donaji Barrera Machuca and Wolfgang Stuerzlinger. “The Effect of Stereo Display Deficiencies on Virtual Hand Pointing”. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI ’19*. the 2019 CHI Conference. Glasgow, Scotland Uk: ACM Press, 2019, pp. 1–14.

[24] Patrick Baudisch, Mike Sinclair, and Andrew Wilson. “Soap: a pointing device that works in mid-air”. In: *Proceedings of the 19th annual ACM symposium on User interface software and technology - UIST ’06*. the 19th annual ACM symposium. Montreux, Switzerland: ACM Press, 2006, p. 43.

[25] Jens Bauer, Sebastian Thelen, and Achim Ebert. “Using smart phones for large-display interaction”. In: *2011 International Conference on User Science and Engineering (i-USEr)*. 2011 International Conference on User Science and Engineering (i-USEr 2011). Selangor, Malaysia: IEEE, Nov. 2011, pp. 42–47.

[26] Alessio Bellino, Federico Cabitza, Giorgio De Michelis, and Flavio De Paoli. “Touch&Screen: widget collection for large screens controlled through smartphones”. In: *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia - MUM ’16*. the 15th International Conference. Rovaniemi, Finland: ACM Press, 2016, pp. 25–35.

[27] Victoria Bellotti, Maribeth Back, W. Keith Edwards, Rebecca E. Grinter, Austin Henderson, and Cristina Lopes. “Making Sense of Sensing Systems: Five Questions for Designers and Researchers”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’02. Minneapolis, Minnesota, USA: Association for Computing Machinery, 2002, pp. 415–422.

[28] Louis-Pierre Bergé, Emmanuel Dubois, and Mathieu Raynal. “Design and Evaluation of an "Around the SmartPhone" Technique for 3D Manipulations on Distant Display”. In: *Proceedings of the 3rd ACM Symposium on Spatial User Interaction - SUI ’15*. the 3rd ACM Symposium. Los Angeles, California, USA: ACM Press, 2015, pp. 69–78.

[29] Louis-Pierre Bergé, Marcos Serrano, Gary Perelman, and Emmanuel Dubois. “Exploring smartphone-based interaction with overview+detail interfaces on 3D public displays”. In: *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services - MobileHCI ’14*. the 16th international conference. Toronto, ON, Canada: ACM Press, 2014, pp. 125–134.

[30] Lonni Besancon, Paul Issartel, Mehdi Ammi, and Tobias Isenberg. “Hybrid Tactile/Tangible Interaction for 3D Data Exploration”. In: *IEEE Transactions on Visualization and Computer Graphics* 23.1 (Jan. 2017), pp. 881–890.

[31] Lonni Besançon, Paul Issartel, Mehdi Ammi, and Tobias Isenberg. “Mouse, Tactile, and Tangible Input for 3D Manipulation”. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI ’17*. the 2017 CHI Conference. Denver, Colorado, USA: ACM Press, 2017, pp. 4727–4740.

[32] Richard A. Bolt. ““Put-that-there”: Voice and gesture at the graphics interface”. In: *Proceedings of the 7th annual conference on Computer graphics and interactive techniques*. SIGGRAPH ’80. Seattle, Washington, USA: Association for Computing Machinery, July 1, 1980, pp. 262–270.

[33] Sebastian Boring, Dominikus Baur, Andreas Butz, Sean Gustafson, and Patrick Baudisch. “Touch projector: mobile interaction through video”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’10. Atlanta, Georgia, USA: Association for Computing Machinery, Apr. 10, 2010, pp. 2287–2296.

[34] Sebastian Boring, Sven Gehring, Alexander Wiethoff, Anna Magdalena Blöckner, Johannes Schöning, and Andreas Butz. “Multi-user interaction on media facades through live video on mobile devices”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’11. Vancouver, BC, Canada: Association for Computing Machinery, May 7, 2011, pp. 2721–2724.

[35] Sebastian Boring, Marko Jurmu, and Andreas Butz. “Scroll, tilt or move it: using mobile phones to continuously control pointers on large public displays”. In: *Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7*. OZCHI ’09. Melbourne, Australia: Association for Computing Machinery, Nov. 23, 2009, pp. 161–168.

[36] Doug Bowman, Sabine Coquillart, Bernd Froehlich, Michitaka Hirose, Yoshifumi Kitamura, Kiyoshi Kiyokawa, and Wolfgang Stuerzlinger. “3D User Interfaces: New Directions and Perspectives”. In: *IEEE computer graphics and applications* 28 (Nov. 2008), pp. 20–36.

[37] Doug A. Bowman and Larry F. Hodges. “An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments”. In: *Proceedings of the 1997 symposium on Interactive 3D graphics*.

I3D '97. Providence, Rhode Island, USA: Association for Computing Machinery, Apr. 30, 1997, 35–ff.

[38] Andrew Bragdon, Rob DeLine, Ken Hinckley, and Meredith Ringel Morris. “Code space: touch + air gesture hybrid interactions for supporting developer meetings”. In: *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*. ITS '11. Kobe, Japan: Association for Computing Machinery, Nov. 13, 2011, pp. 212–221.

[39] Frederik Brudy, Christian Holz, Roman Rädle, Chi-Jui Wu, Steven Houben, Clemens Nylandsted Klokmose, and Nicolai Marquardt. “Cross-Device Taxonomy: Survey, Opportunities and Challenges of Interactions Spanning Across Multiple Devices”. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. Glasgow, Scotland Uk: Association for Computing Machinery, May 2, 2019, pp. 1–28.

[40] Wolfgang Büschel, Annett Mitschick, Thomas Meyer, and Raimund Dachselt. “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, Oct. 1, 2019, pp. 1–13.

[41] Wolfgang Büschel, Patrick Reipschläger, and Raimund Dachselt. “Foldable3D: Interacting with 3D Content Using Dual-Display Devices”. In: *Proceedings of the 2016 ACM on Interactive Surfaces and Spaces - ISS '16*. the 2016 ACM. Niagara Falls, Ontario, Canada: ACM Press, 2016, pp. 367–372.

[42] Wolfgang Büschel, Patrick Reipschläger, Ricardo Langner, and Raimund Dachselt. “Investigating the Use of Spatial Interaction for 3D Data Visualization on Mobile Devices”. In: *Proceedings of the Interactive Surfaces and Spaces on ZZZ - ISS '17*. the Interactive Surfaces and Spaces. Brighton, United Kingdom: ACM Press, 2017, pp. 62–71.

[43] Llogari Casas, Loïc Ciccone, Gökçen Çimen, Pablo Wiedemann, Matthias Fauconneau, Robert W. Sumner, and Kenny Mitchell. “Multi-reality games: an experience across the entire reality-virtuality continuum”. In: *Proceedings of the 16th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry - VRCAI '18*. the 16th

ACM SIGGRAPH International Conference. Tokyo, Japan: ACM Press, 2018, pp. 1–4.

[44] Géry Casiez, Nicolas Roussel, and Daniel Vogel. “1 € filter: a simple speed-based low-pass filter for noisy input in interactive systems”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’12. Austin, Texas, USA: Association for Computing Machinery, May 5, 2012, pp. 2527–2530.

[45] Géry Casiez, Nicolas Roussel, and Daniel Vogel. *1€ Filter*. Jan. 5, 2020.

[46] Elie Cattan, Amélie Rochet-Capellan, Pascal Perrier, and François Bérard. “Does Practice Make Perfect?” In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. CHI ’17: CHI Conference on Human Factors in Computing Systems. Denver Colorado USA: ACM, May 2, 2017, pp. 5619–5629.

[47] Ishan Chatterjee, Robert Xiao, and Chris Harrison. “Gaze+Gesture: Expressive, Precise and Targeted Free-Space Interactions”. In: *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction - ICMI ’15*. the 2015 ACM. Seattle, Washington, USA: ACM Press, 2015, pp. 131–138.

[48] Xiang ’Anthony’ Chen, Tovi Grossman, Daniel J. Wigdor, and George Fitzmaurice. “Duet: exploring joint interactions on a smart phone and a smart watch”. In: *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI ’14*. the 32nd annual ACM conference. Toronto, Ontario, Canada: ACM Press, 2014, pp. 159–168.

[49] Xiang ’Anthony’ Chen, Nicolai Marquardt, Anthony Tang, Sebastian Boring, and Saul Greenberg. “Extending a mobile device’s interaction space through body-centric interaction”. In: *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services - MobileHCI ’12*. the 14th international conference. San Francisco, California, USA: ACM Press, 2012, p. 151.

[50] Yuan Chen, Keiko Katsuragawa, and Edward Lank. “Understanding Viewport- and World-based Pointing with Everyday Smart Devices in Immersive Augmented Reality”. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI ’20. Honolulu, HI, USA: Association for Computing Machinery, Apr. 21, 2020, pp. 1–13.

[51] Muratcan Cicek, Jinrong Xie, Qiaosong Wang, and Robinson Piramuthu. "Mobile Head Tracking for eCommerce and Beyond". In: *arXiv:1812.07143 [cs]* (Dec. 17, 2018). arXiv: 1812.07143.

[52] Christopher Clarke and Hans Gellersen. "MatchPoint: Spontaneous Spatial Coupling of Body Movement for Touchless Pointing". In: *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology - UIST '17*. the 30th Annual ACM Symposium. Quebec City, QC, Canada: ACM Press, 2017, pp. 179–192.

[53] Aurélie Cohé, Fabrice Dècle, and Martin Hachet. "tBox: a 3d transformation widget designed for touch-screens". In: *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*. the 2011 annual conference. Vancouver, BC, Canada: ACM Press, 2011, p. 3005.

[54] Klen Čopič Pucihar, Paul Coulton, and Jason Alexander. "The use of surrounding visual context in handheld AR: device vs. user perspective rendering". In: *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*. the 32nd annual ACM conference. Toronto, Ontario, Canada: ACM Press, 2014, pp. 197–206.

[55] Klen Čopič Pucihar, Matjaž Kljun, and Paul Coulton. "Using a Mobile Phone as a 2D Virtual Tracing Tool: Static Peephole vs. Magic Lens". In: *Technology and Intimacy: Choice or Coercion*. Ed. by David Kreps, Gordon Fletcher, and Marie Griffiths. Vol. 474. Series Title: IFIP Advances in Information and Communication Technology. Cham: Springer International Publishing, 2016, pp. 277–288.

[56] Emmanuel Courtoux, Caroline Appert, and Olivier Chapuis. "SurfAirs: Surface and Mid-Air Input for Large Vertical Displays". In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. CHI '23. Hamburg, Germany: Association for Computing Machinery, 2023.

[57] Carolina Cruz-Neira, Daniel J. Sandin, Thomas A. DeFanti, Robert V. Kenyon, and John C. Hart. "The CAVE: audio visual experience automatic virtual environment". In: *Communications of the ACM* 35.6 (June 1, 1992), pp. 64–72.

[58] Raimund Dachselt and Robert Buchholz. "Natural throw and tilt interaction between mobile phones and distant displays". In: *Proceedings of the 27th international conference extended abstracts on Human factors in com-*

puting systems - CHI EA '09. the 27th international conference extended abstracts. Boston, MA, USA: ACM Press, 2009, p. 3253.

[59] Florian Daiber, Marco Speicher, Sven Gehring, Markus Löchtefeld, and Antonio Krüger. "Interacting with 3D Content on Stereoscopic Displays". In: *Proceedings of The International Symposium on Pervasive Displays*. PerDis '14. Copenhagen, Denmark: Association for Computing Machinery, June 3, 2014, pp. 32–37.

[60] H. Debarba, L. Nedel, and A. Maciel. "LOP-cursor: Fast and precise interaction with tiled displays using one hand and levels of precision". In: *2012 IEEE Symposium on 3D User Interfaces (3DUI)*. 2012 IEEE Symposium on 3D User Interfaces (3DUI). Costa Mesa, CA: IEEE, Mar. 2012, pp. 125–132.

[61] Henrique G. Debarba, Jerônimo G. Grandi, Anderson Maciel, Luciana Nedel, and Ronan Boulic. "Disambiguation Canvas: A Precise Selection Technique for Virtual Environments". In: *Human-Computer Interaction – INTERACT 2013*. Ed. by Paula Kotzé, Gary Marsden, Gitte Lindgaard, Janet Wesson, and Marco Winckler. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 388–405.

[62] Yu Ding, Yuting Zhang, Meihua Xiao, and Zhigang Deng. "A Multifaceted Study on Eye Contact based Speaker Identification in Three-party Conversations". In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. CHI '17: CHI Conference on Human Factors in Computing Systems. Denver Colorado USA: ACM, May 2, 2017, pp. 3011–3021.

[63] Alan Dix, Janet Finley, Gregory Abowd, and Russell Beale. *Human-Computer Interaction (2nd Ed.)* USA: Prentice-Hall, Inc., 1998.

[64] Yuan Du, Haoyi Ren, Gang Pan, and Shjian Li. "Tilt & touch: mobile phone for 3D interaction". In: (2011), p. 2.

[65] Hugh Durrant-Whyte and Tim Bailey. "Simultaneous Localisation and Mapping (SLAM): Part I The Essential Algorithms". In: (Jan. 1, 2006), p. 9.

[66] Yutaka Endo, Dai Fujita, and Takashi Komuro. "Distant Pointing User Interfaces based on 3D Hand Pointing Recognition". In: *Proceedings of the Interactive Surfaces and Spaces on ZZZ - ISS '17*. the Interactive Surfaces and Spaces. Brighton, United Kingdom: ACM Press, 2017, pp. 413–416.

[67] Daniel Fallman. “Design-Oriented Human-Computer Interaction”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’03. Ft. Lauderdale, Florida, USA: Association for Computing Machinery, 2003, pp. 225–232.

[68] Mingming Fan, Donald Patterson, and Yuanchun Shi. “When camera meets accelerometer: a novel way for 3d interaction of mobile phone”. In: (2012), p. 6.

[69] Matthias Finke, Nima Kaviani, Ivy Wang, Vincent Tsao, Sidney Fels, and Rodger Lea. “Investigating distributed user interfaces across interactive large displays and mobile devices”. In: *Proceedings of the International Conference on Advanced Visual Interfaces - AVI ’10*. the International Conference. Roma, Italy: ACM Press, 2010, p. 413.

[70] George Fitzmaurice and William Buxton. “The Chameleon: spatially aware palmtop computers”. In: *Conference Companion on Human Factors in Computing Systems*. CHI ’94. Boston, Massachusetts, USA: Association for Computing Machinery, Apr. 28, 1994, pp. 451–452.

[71] James D. Foley, Victor L. Wallace, and Peggy Chan. “The human factors of computer graphics interaction techniques”. In: *IEEE Computer Graphics and Applications* 4.11 (1984), pp. 13–48.

[72] Clifton Forlines, Daniel Vogel, and Ravin Balakrishnan. “HybridPointing: fluid switching between absolute and relative pointing with a direct input device”. In: *Proceedings of the 19th annual ACM symposium on User interface software and technology - UIST ’06*. the 19th annual ACM symposium. Montreux, Switzerland: ACM Press, 2006, p. 211.

[73] Jérémie Francone and Laurence Nigay. “Using the user’s point of view for interaction on mobile devices”. In: *23rd French Speaking Conference on Human-Computer Interaction - IHM ’11*. 23rd French Speaking Conference. Sophia Antipolis, France: ACM Press, 2011, p. 1.

[74] Jorge Fuentes-Pacheco, José Ruiz-Ascencio, and Juan Manuel Rendón-Mancha. “Visual simultaneous localization and mapping: a survey”. In: *Artificial Intelligence Review* 43.1 (Jan. 2015), pp. 55–81.

[75] Leo Gombač, Klen Čopič Pucihar, Matjaž Kljun, Paul Coulton, and Jan Grbac. “3D Virtual Tracing and Depth Perception Problem on Mobile AR”. In: *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA ’16*. the 2016 CHI Conference Extended Abstracts. San Jose, California, USA: ACM Press, 2016, pp. 1849–1856.

[76] Google. *ARCore*. URL: %5Curl%7Bhttps://developers.google.com/ar/%7D.

[77] Google. *ARCore - World Tracking*. URL: https://developers.google.com/ar/discover/concepts%5C#motion_tracking.

[78] Daniel Gotsch, Xujing Zhang, Jesse Burstyn, and Roel Vertegaal. “HoloFlex: A Flexible Holographic Smartphone with Bend Input”. In: *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA ’16*. the 2016 CHI Conference Extended Abstracts. San Jose, California, USA: ACM Press, 2016, pp. 3675–3678.

[79] Henning Graf and Dr Klaus Jung. “The Smartphone as a 3D Input Device”. In: (Jan. 1, 2012), p. 4.

[80] Jerônimo Gustavo Grandi, Henrique Galvan Debarba, Luciana Nedel, and Anderson Maciel. “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*. the 2017 CHI Conference. Denver, Colorado, USA: ACM Press, 2017, pp. 5881–5891.

[81] Nielsen Norman Group. *Inclusive Design*. URL: https://www.nngroup.com/articles/inclusive-design/.

[82] Aakar Gupta, Muhammed Anwar, and Ravin Balakrishnan. “Porous Interfaces for Small Screen Multitasking using Finger Identification”. In: *Proceedings of the 29th Annual Symposium on User Interface Software and Technology - UIST ’16*. the 29th Annual Symposium. Tokyo, Japan: ACM Press, 2016, pp. 145–156.

[83] Martin Hachet, Joachim Pouderoux, and Pascal Guitton. “A camera-based interface for interaction with mobile handheld computers”. In: *Proceedings of the 2005 symposium on Interactive 3D graphics and games - SI3D ’05*. the 2005 symposium. Washington, District of Columbia: ACM Press, 2005, p. 65.

[84] Jonna R. Häkkilä, Maaret Posti, Stefan Schneegass, Florian Alt, Kunter Gultekin, and Albrecht Schmidt. “Let me catch this! experiencing interactive 3D cinema through collecting content with a mobile phone”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’14. Toronto, Ontario, Canada: Association for Computing Machinery, Apr. 26, 2014, pp. 1011–1020.

[85] Jari Hannuksela, Pekka Sangi, and Janne Heikkilä. “Vision-based motion estimation for interaction with mobile devices”. In: *Computer Vision and Image Understanding* 108.1 (Oct. 2007), pp. 188–195.

[86] Faizan Haque, Mathieu Nancel, and Daniel Vogel. “Myopoint: Pointing and Clicking Using Forearm Mounted Electromyography and Inertial Motion Sensors”. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI ’15*. the 33rd Annual ACM Conference. Seoul, Republic of Korea: ACM Press, 2015, pp. 3653–3656.

[87] Robert Hardy and Enrico Rukzio. “Touch & interact: touch-based interaction of mobile phones with displays”. In: *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services - MobileHCI ’08*. the 10th international conference. Amsterdam, The Netherlands: ACM Press, 2008, p. 245.

[88] Chris Harrison, Hrvoje Benko, and Andrew D. Wilson. “OmniTouch: wearable multitouch interaction everywhere”. In: *Proceedings of the 24th annual ACM symposium on User interface software and technology - UIST ’11*. the 24th annual ACM symposium. Santa Barbara, California, USA: ACM Press, 2011, p. 441.

[89] Jeremy Hartmann, Christian Holz, Eyal Ofek, and Andrew D Wilson. “RealityCheck: Blending Virtual Environments with Situated Physical Reality”. In: (2019), p. 12.

[90] Jeremy Hartmann and Daniel Vogel. “An Evaluation of Mobile Phone Pointing in Spatial Augmented Reality”. In: *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI EA ’18. Montreal QC, Canada: Association for Computing Machinery, Apr. 20, 2018, pp. 1–6.

[91] Khalad Hasan, David Ahlström, and Pourang Irani. “SAMMI: A Spatially-Aware Multi-Mobile Interface for Analytic Map Navigation Tasks”. In: *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI ’15*. the 17th International Conference. Copenhagen, Denmark: ACM Press, 2015, pp. 36–45.

[92] Khalad Hasan, David Ahlström, Junhyeok Kim, and Pourang Irani. “AirPanes: Two-Handed Around-Device Interaction for Pane Switching on Smartphones”. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. CHI ’17: CHI Conference on Human Factors in Computing Systems. Denver Colorado USA: ACM, May 2, 2017, pp. 679–691.

[93] Jay Henderson, Sachi Mizobuchi, Wei Li, and Edward Lank. “Exploring Cross-Modal Training via Touch to Learn a Mid-Air Marking Menu Gesture Set”. In: *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services*. MobileHCI ’19. Taipei, Taiwan: Association for Computing Machinery, Oct. 1, 2019, pp. 1–9.

[94] A. Henrysson, M. Billinghurst, and M. Ollila. “Face to face collaborative AR on mobile phones”. In: *Fourth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR’05)*. Fourth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR’05). Vienna, Austria: IEEE, 2005, pp. 80–89.

[95] Anders Henrysson and Mark Billinghurst. “Using a mobile phone for 6 DOF mesh editing”. In: *Proceedings of the 8th ACM SIGCHI New Zealand chapter’s international conference on Computer-human interaction: design centered HCI*. CHINZ ’07. Hamilton, New Zealand: Association for Computing Machinery, July 2, 2007, pp. 9–16.

[96] Anders Henrysson, Mark Billinghurst, and Mark Ollila. “Virtual object manipulation using a mobile phone”. In: *Proceedings of the 2005 international conference on Augmented tele-existence*. ICAT ’05. Christchurch, New Zealand: Association for Computing Machinery, Dec. 5, 2005, pp. 164–171.

[97] Jacob Schiefer Alex Hill. “Virtual Transparency: Introducing Parallax View into Video See-through AR”. In: (2011), p. 2.

[98] Juan David Hincapié-Ramos, Xiang Guo, Paymahn Moghadasian, and Pourang Irani. “Consumed Endurance: A Metric to Quantify Arm Fatigue of Mid-air Interactions”. In: *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*. CHI ’14. Toronto, Ontario, Canada: ACM, 2014, pp. 1063–1072.

[99] Ken Hinckley, Randy Pausch, John C. Goble, and Neal F. Kassell. “A survey of design issues in spatial input”. In: *Proceedings of the 7th annual ACM symposium on User interface software and technology*. UIST ’94. Marina del Rey, California, USA: Association for Computing Machinery, Nov. 2, 1994, pp. 213–222.

[100] Ken Hinckley, Jeff Pierce, Mike Sinclair, and Eric Horvitz. “Sensing techniques for mobile interaction”. In: *Proceedings of the 13th annual ACM symposium on User interface software and technology - UIST ’00*. the 13th annual ACM symposium. San Diego, California, United States: ACM Press, 2000, pp. 91–100.

[101] Tom Horak, Sriram Karthik Badam, Niklas Elmqvist, and Raimund Dachselt. “When David Meets Goliath: Combining Smartwatches with a Large Vertical Display for Visual Data Exploration”. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI ’18*. the 2018 CHI Conference. Montreal QC, Canada: ACM Press, 2018, pp. 1–13.

[102] Steven Houben and Nicolai Marquardt. “WatchConnect: A Toolkit for Prototyping Smartwatch-Centric Cross-Device Applications”. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI ’15. Seoul, Republic of Korea: Association for Computing Machinery, Apr. 18, 2015, pp. 1247–1256.

[103] Stephanie Houde. “Iterative design of an interface for easy 3-D direct manipulation”. In: *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI ’92*. the SIGCHI conference. Monterey, California, United States: ACM Press, 1992, pp. 135–142.

[104] Yi-Ta Hsieh, Antti Jylhä, Valeria Orso, Luciano Gamberini, and Giulio Jacucci. “Designing a Willing-to-Use-in-Public Hand Gestural Interaction Technique for Smart Glasses”. 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. 4203–4215.

[105] Hilary Hutchinson, Wendy Mackay, Bo Westerlund, Benjamin B. Bederson, Allison Druin, Catherine Plaisant, Michel Beaudouin-Lafon, Stéphane Conversy, Helen Evans, Heiko Hansen, Nicolas Roussel, and Björn Eiderbäck. “Technology Probes: Inspiring Design for and with Families”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '03. Ft. Lauderdale, Florida, USA: Association for Computing Machinery, 2003, pp. 17–24.

[106] *Ergonomic requirements for office work with visual display terminals (VDTs) - Part 9: Requirements for non-keyboard input devices*. Standard. Geneva, CH: International Organization for Standardization, 2000.

[107] Mohit Jain, Andy Cockburn, and Sriganesh Madhvanath. “Comparison of Phone-Based Distal Pointing Techniques for Point-Select Tasks”. In: *Human-Computer Interaction – INTERACT 2013*. Ed. by Paula Kotzé, Gary Marsden, Gitte Lindgaard, Janet Wesson, and Marco Winckler. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2013, pp. 714–721.

[108] Mikkel R. Jakobsen and Kasper Hornbæk. “Is Moving Improving?: Some Effects of Locomotion in Wall-Display Interaction”. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*. the 33rd Annual ACM Conference. Seoul, Republic of Korea: ACM Press, 2015, pp. 4169–4178.

[109] Mikkel R. Jakobsen, Yvonne Jansen, Sebastian Boring, and Kasper Hornbæk. “Should I Stay or Should I Go? Selecting Between Touch and Mid-Air Gestures for Large-Display Interaction”. In: *15th IFIP TC 13 International Conference on Human-Computer Interaction — INTERACT 2015 - Volume 9298*. Berlin, Heidelberg: Springer-Verlag, Sept. 14, 2015, pp. 455–473.

[110] Shahram Jalaliniya, Diako Mardanbegi, and Thomas Pederson. “MAGIC pointing for eyewear computers”. In: *Proceedings of the 2015 ACM International Symposium on Wearable Computers*. ISWC '15. Osaka, Japan: Association for Computing Machinery, Sept. 7, 2015, pp. 155–158.

[111] Sujin Jang, Wolfgang Stuerzlinger, Satyajit Ambike, and Karthik Ramani. “Modeling Cumulative Arm Fatigue in Mid-Air Interaction Based on Perceived Exertion and Kinetics of Arm Motion”. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. CHI ’17. Denver, Colorado, USA: ACM, 2017, pp. 3328–3339.

[112] Florian Jann, Sascha Reinhold, and Michael Teistler. “Nutzung eines Smartphones als virtuelle Sonde im medizinischen Ultraschalltraining: Six-Degrees-of-Freedom-Tracking mittels ARCore”. In: *Proceedings of Mensch und Computer 2019*. MuC’19. Hamburg, Germany: Association for Computing Machinery, Sept. 8, 2019, pp. 759–763.

[113] Yvonne Jansen, Jonas Schjerlund, and Kasper Hornbæk. “Effects of Locomotion and Visual Overview on Spatial Memory when Interacting with Wall Displays”. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI ’19*. the 2019 CHI Conference. Glasgow, Scotland Uk: ACM Press, 2019, pp. 1–12.

[114] Seokhee Jeon, Jane Hwang, Gerard J. Kim, and Mark Billinghurst. *Interaction with large ubiquitous displays using camera-equipped mobile phones*. Feb. 1, 2010. URL: https://doi.org/10.1007/s00779-009-0249-0.

[115] Hao Jiang, Eyal Ofek, Neema Moraveji, and Yuanchun Shi. “Direct pointer: direct manipulation for large-display interaction using handheld cameras”. In: *Proceedings of the SIGCHI conference on Human Factors in computing systems - CHI ’06*. the SIGCHI conference. Montreal, Quebec, Canada: ACM Press, 2006, p. 1107.

[116] Zhiping Jiang, Jinsong Han, Chen Qian, Wei Xi, Kun Zhao, Han Ding, Shaojie Tang, Jizhong Zhao, and Panlong Yang. “VADS: Visual attention detection with a smartphone”. In: *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*. IEEE INFOCOM 2016 - IEEE Conference on Computer Communications. San Francisco, CA, USA: IEEE, Apr. 2016, pp. 1–9.

[117] Keiko Katsuragawa, Krzysztof Pietroszek, James R. Wallace, and Edward Lank. “Watchpoint: Freehand Pointing with a Smartwatch in a Ubiquitous Display Environment”. In: *Proceedings of the International Working Conference on Advanced Visual Interfaces - AVI ’16*. the International Working Conference. Bari, Italy: ACM Press, 2016, pp. 128–135.

[118] Nicholas Katzakis and Masahiro Hori. “Mobile Phones as 3-DOF Controllers: A Comparative Study”. In: *2009 Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing*. 2009 International Conference on Dependable, Autonomic and Secure Computing (DASC). Chengdu, China: IEEE, Dec. 2009, pp. 345–349.

[119] Nicholas Katzakis, Kiyoshi Kiyokawa, and Haruo Takemura. “Plane-casting: 3D cursor control with a SmartPhone”. In: *Proceedings of the 11th Asia Pacific Conference on Computer Human Interaction - APCHI ’13*. the 11th Asia Pacific Conference. Bangalore, India: ACM Press, 2013, pp. 199–200.

[120] Nicholas Katzakis, Robert J Teather, Kiyoshi Kiyokawa, and Haruo Takemura. “INSPECT: extending plane-casting for 6-DOF control”. In: *Human-centric Computing and Information Sciences* 5.1 (Dec. 2015), p. 22.

[121] Bonifaz Kaufmann and David Ahlström. “Studying spatial memory and map navigation performance on projector phones with peephole interaction”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI ’13*. the SIGCHI Conference. Paris, France: ACM Press, 2013, p. 3173.

[122] Nima Kaviani, Matthias Finke, Sidney Fels, Rodger Lea, and Hua Wang. “What goes where?: designing interactive large public display applications for mobile device interaction”. In: *Proceedings of the First International Conference on Internet Multimedia Computing and Service - ICIMCS ’09*. the First International Conference. Kunming, Yunnan, China: ACM Press, 2009, p. 129.

[123] Juha Kela, Panu Korpipää, Jani Mäntyjärvi, Sanna Kallio, Giuseppe Savino, Luca Jozzo, and Sergio Di Marca. “Accelerometer-based gesture control for a design environment”. In: *Personal and Ubiquitous Computing* 10.5 (Aug. 2006), pp. 285–299.

[124] Mohamed Khamis, Florian Alt, and Andreas Bulling. “Challenges and design space of gaze-enabled public displays”. In: *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct - UbiComp ’16*. the 2016 ACM International Joint Conference. Heidelberg, Germany: ACM Press, 2016, pp. 1736–1745.

[125] Mohamed Khamis, Florian Alt, and Andreas Bulling. “The past, present, and future of gaze-enabled handheld mobile devices: survey and lessons learned”. In: *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI ’18*. the 20th International Conference. Barcelona, Spain: ACM Press, 2018, pp. 1–17.

[126] Mohamed Khamis, Anita Baier, Niels Henze, Florian Alt, and Andreas Bulling. “Understanding Face and Eye Visibility in Front-Facing Cameras of Smartphones used in the Wild”. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI ’18*. the 2018 CHI Conference. Montreal QC, Canada: ACM Press, 2018, pp. 1–12.

[127] Mohamed Khamis, Axel Hoesl, Alexander Klimczak, Martin Reiss, Florian Alt, and Andreas Bulling. “EyeScout: Active Eye Tracking for Position and Movement Independent Gaze Interaction with Large Public Displays”. In: *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*. UIST ’17. Québec City, QC, Canada: Association for Computing Machinery, Oct. 20, 2017, pp. 155–166.

[128] Mohamed Khamis, Anna Kienle, Florian Alt, and Andreas Bulling. “GazeDrone: Mobile Eye-Based Interaction in Public Space Without Augmenting the User”. In: *Proceedings of the 4th ACM Workshop on Micro Aerial Vehicle Networks, Systems, and Applications - DroNet’18*. the 4th ACM Workshop. Munich, Germany: ACM Press, 2018, pp. 66–71.

[129] Azam Khan, George Fitzmaurice, Don Almeida, Nicolas Burtnyk, and Gordon Kurtenbach. “A remote control interface for large displays”. In: *Proceedings of the 17th annual ACM symposium on User interface software and technology - UIST ’04*. the 17th annual ACM symposium. Santa Fe, NM, USA: ACM Press, 2004, p. 127.

[130] U. Kister, K. Klamka, C. Tominski, and R. Dachselt. “GraSp : Combining Spatially-aware Mobile Devices and a Display Wall for Graph Visualization and Interaction”. In: *Computer Graphics Forum* 36.3 (June 2017), pp. 503–514.

[131] Ulrike Kister, Patrick Reipschläger, Fabrice Matulic, and Raimund Dachselt. “BodyLenses: Embodied Magic Lenses and Personal Territories for Wall Displays”. In: *Proceedings of the 2015 International Conference on Inter-*

active Tabletops & Surfaces - ITS '15. the 2015 International Conference. Madeira, Portugal: ACM Press, 2015, pp. 117–126.

[132] Werner A. König, Jens Gerken, Stefan Dierdorf, and Harald Reiterer. "Adaptive pointing: implicit gain adaptation for absolute pointing devices". In: *CHI '09 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '09. Boston, MA, USA: Association for Computing Machinery, Apr. 4, 2009, pp. 4171–4176.

[133] Robert Kooima. "Generalized Perspective Projection". In: (Jan. 1, 2005), p. 7.

[134] Kyle Krafka, Aditya Khosla, Petr Kellnhofer, Harini Kannan, Suchendra Bhandarkar, Wojciech Matusik, and Antonio Torralba. "Eye Tracking for Everyone". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, June 2016, pp. 2176–2184.

[135] Christian Kray, Daniel Nesbitt, John Dawson, and Michael Rohs. "User-defined gestures for connecting mobile phones, public displays, and tabletops". In: *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*. MobileHCI '10. Lisbon, Portugal: Association for Computing Machinery, Sept. 7, 2010, pp. 239–248.

[136] Myron W. Krueger, Thomas Gionfriddo, and Katrin Hinrichsen. "VIDEOPLACE—an artificial reality". In: *ACM SIGCHI Bulletin* 16.4 (Apr. 1, 1985), pp. 35–40.

[137] Ernst Kruijff, J. Edward Swan, and Steven Feiner. "Perceptual issues in augmented reality revisited". In: *2010 IEEE International Symposium on Mixed and Augmented Reality*. 2010 9th IEEE International Symposium on Mixed and Augmented Reality (ISMAR). Seoul, Korea (South): IEEE, Oct. 2010, pp. 3–12.

[138] Mikko Kytö, Barrett Ens, Thammathip Piumsomboon, Gun A. Lee, and Mark Billinghurst. "Pinpointing: Precise Head- and Eye-Based Target Selection for Augmented Reality". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. the 2018 CHI Conference. Montreal QC, Canada: ACM Press, 2018, pp. 1–14.

[139] David Lakatos, Matthew Blackshaw, Alex Olwal, Zachary Barryte, Ken Perlin, and Hiroshi Ishii. “T(ether): spatially-aware handhelds, gestures and proprioception for multi-user 3D modeling and animation”. In: *Proceedings of the 2nd ACM symposium on Spatial user interaction - SUI ’14*. the 2nd ACM symposium. Honolulu, Hawaii, USA: ACM Press, 2014, pp. 90–93.

[140] Christian Lander, Sven Gehring, Antonio Krüger, Sebastian Boring, and Andreas Bulling. “GazeProjector: Accurate Gaze Estimation and Seamless Gaze Interaction Across Multiple Displays”. In: *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology - UIST ’15*. the 28th Annual ACM Symposium. Daegu, Kyungpook, Republic of Korea: ACM Press, 2015, pp. 395–404.

[141] Ricardo Langner and Raimund Dachselt. “Towards Visual Data Exploration at Wall-Sized Displays by Combining Physical Navigation with Spatially-Aware Devices”. In: (2018), p. 2.

[142] Ricardo Langner, Ulrike Kister, and Raimund Dachselt. “Multiple Coordinated Views at Large Displays for Multiple Users: Empirical Findings on User Behavior, Movements, and Distances”. In: *IEEE Transactions on Visualization and Computer Graphics* 25.1 (Jan. 2019), pp. 608–618.

[143] Ricardo Langner, Ulrike Kister, Marc Satkowski, and Raimund Dachselt. “Combining Interactive Large Displays and Smartphones to Enable Data Analysis from Varying Distances”. In: (2018), p. 4.

[144] Joseph J. LaViola, Ernst Kruijff, Ryan P. McMahan, Doug A. Bowman, and Ivan Poupyrev. *3D User Interfaces: Theory and Practice*. Addison-Wesley, 2017. 591 pp.

[145] Huy Viet Le, Sven Mayer, and Niels Henze. “InfiniTouch: Finger-Aware Interaction on Fully Touch Sensitive Smartphones”. In: *The 31st Annual ACM Symposium on User Interface Software and Technology - UIST ’18*. The 31st Annual ACM Symposium. Berlin, Germany: ACM Press, 2018, pp. 779–792.

[146] Johnny Chung Lee. “Hacking the Nintendo Wii Remote”. In: *IEEE Pervasive Computing* 7.3 (Jan. 7, 2008). Conference Name: IEEE Pervasive Computing, pp. 39–45.

[147] Jong-In Lee, Sunjun Kim, Masaaki Fukumoto, and Byungjoo Lee. “Reflector: Distance-Independent, Private Pointing on a Reflective Screen”. In: *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology - UIST ’17*. the 30th Annual ACM Symposium. Quebec City, QC, Canada: ACM Press, 2017, pp. 351–364.

[148] Sang-won Leigh, Philipp Schoessler, Felix Heibeck, Pattie Maes, and Hiroshi Ishii. “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 ’14*. the Ninth International Conference. Stanford, California, USA: ACM Press, 2015, pp. 89–96.

[149] Hai-Ning Liang, Yuwei Shi, Feiyu Lu, Jizhou Yang, and Konstantinos Papangelis. “VRMController: an input device for navigation activities in virtual reality environments”. In: *Proceedings of the 15th ACM SIGGRAPH Conference on Virtual-Reality Continuum and Its Applications in Industry - VRCAI ’16*. the 15th ACM SIGGRAPH Conference. Zhuhai, China: ACM Press, 2016, pp. 455–460.

[150] Can Liu, Olivier Chapuis, Michel Beaudouin-Lafon, and Eric Lecolinet. “CoReach: Cooperative Gestures for Data Manipulation on Wall-sized Displays”. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. CHI ’17. Denver, Colorado, USA: Association for Computing Machinery, May 2, 2017, pp. 6730–6741.

[151] Can Liu, Olivier Chapuis, Michel Beaudouin-Lafon, Eric Lecolinet, and Wendy E. Mackay. “Effects of display size and navigation type on a classification task”. In: *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI ’14*. the 32nd annual ACM conference. Toronto, Ontario, Canada: ACM Press, 2014, pp. 4147–4156.

[152] Mingyu Liu, Mathieu Nancel, and Daniel Vogel. “Gunslinger: Subtle Arms-down Mid-air Interaction”. In: *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology - UIST ’15*. the 28th Annual ACM Symposium. Daegu, Kyungpook, Republic of Korea: ACM Press, 2015, pp. 63–71.

[153] David Lopez, Lora Oehlberg, Candemir Doger, and Tobias Isenberg. “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 (May 1, 2016), pp. 1616–1629.

[154] Xiaolong Lou, Andol X. Li, Ren Peng, and Preben Hansen. "Optimising Free Hand Selection in Large Displays by Adapting to User's Physical Movements". In: *Proceedings of the 2016 Symposium on Spatial User Interaction*. SUI '16. Tokyo, Japan: Association for Computing Machinery, Oct. 15, 2016, pp. 23–31.

[155] Zhihan Lv, Shengzhong Feng, Muhammad Sikandar Lal Khan, Shafiq Ur Réhman, and Haibo Li. "Foot motion sensing: augmented game interface based on foot interaction for smartphone". In: *Proceedings of the extended abstracts of the 32nd annual ACM conference on Human factors in computing systems - CHI EA '14*. the extended abstracts of the 32nd annual ACM conference. Toronto, Ontario, Canada: ACM Press, 2014, pp. 293–296.

[156] I. Scott MacKenzie, Tatu Kauppinen, and Miika Silfverberg. "Accuracy Measures for Evaluating Computer Pointing Devices". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '01. Seattle, Washington, USA: ACM, 2001, pp. 9–16.

[157] Scott MacKenzie. *Fitts' Law Software*. URL: http://www.yorku.ca/mack/FittsLawSoftware.

[158] Ville Mäkelä, Mohamed Khamis, Lukas Mecke, Jobin James, Markku Turunen, and Florian Alt. "Pocket Transfers: Interaction Techniques for Transferring Content from Situated Displays to Mobile Devices". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18. Montreal QC, Canada: Association for Computing Machinery, Apr. 19, 2018, pp. 1–13.

[159] Alexander Mariakakis, Mayank Goel, Md Tanvir Islam Aumi, Shwetak N. Patel, and Jacob O. Wobbrock. "SwitchBack: Using Focus and Saccade Tracking to Guide Users' Attention for Mobile Task Resumption". In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*. the 33rd Annual ACM Conference. Seoul, Republic of Korea: ACM Press, 2015, pp. 2953–2962.

[160] Anders Markussen, Sebastian Boring, Mikkel R. Jakobsen, and Kasper Hornbæk. "Off-Limits: Interacting Beyond the Boundaries of Large Displays". In: *Proceedings of the 2016 CHI Conference on Human Factors in*

Computing Systems. CHI'16: CHI Conference on Human Factors in Computing Systems. San Jose California USA: ACM, May 7, 2016, pp. 5862–5873.

[161] Nicolai Marquardt, Robert Diaz-Marino, Sebastian Boring, and Saul Greenberg. “The proximity toolkit: prototyping proxemic interactions in ubiquitous computing ecologies”. In: *Proceedings of the 24th annual ACM symposium on User interface software and technology - UIST '11*. the 24th annual ACM symposium. Santa Barbara, California, USA: ACM Press, 2011, p. 315.

[162] Nicolai Marquardt, Ken Hinckley, and Saul Greenberg. “Cross-device interaction via micro-mobility and f-formations”. In: *Proceedings of the 25th annual ACM symposium on User interface software and technology*. UIST '12. Cambridge, Massachusetts, USA: Association for Computing Machinery, Oct. 7, 2012, pp. 13–22.

[163] Anthony Martinet, Géry Casiez, and Laurent Grisoni. “The effect of DOF separation in 3D manipulation tasks with multi-touch displays”. In: *Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology - VRST '10*. the 17th ACM Symposium. Hong Kong: ACM Press, 2010, p. 111.

[164] Asier Marzo, Benoît Bossavit, and Martin Hachet. “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*. the 2nd ACM symposium. Honolulu, Hawaii, USA: ACM Press, 2014, pp. 13–16.

[165] Y. Matsuda, F. Shibata, A. Kimura, and H. Tamura. “Poster: Creating a user-specific perspective view for mobile mixed reality systems on smartphones”. In: *2013 IEEE Symposium on 3D User Interfaces (3DUI)*. Mar. 2013, pp. 157–158.

[166] Yoshio Matsuda and Takashi Komuro. “Dynamic layout optimization for multi-user interaction with a large display”. In: *Proceedings of the 25th International Conference on Intelligent User Interfaces*. IUI '20. Cagliari, Italy: Association for Computing Machinery, Mar. 17, 2020, pp. 401–409.

[167] Cameron Appel Max Roser and Hannah Ritchie. “Human Height”. In: ().

[168] Sven Mayer, Gierad Laput, and Chris Harrison. “Enhancing Mobile Voice Assistants with WorldGaze”. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI ’20. Honolulu, HI, USA: Association for Computing Machinery, Apr. 21, 2020, pp. 1–10.

[169] Sven Mayer, Lars Lischke, Jens Emil Grønbæk, Zhanna Sarsenbayeva, Jonas Vogelsang, Paweł W. Woźniak, Niels Henze, and Giulio Jacucci. “Pac-Many: Movement Behavior when Playing Collaborative and Competitive Games on Large Displays”. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI ’18*. the 2018 CHI Conference. Montreal QC, Canada: ACM Press, 2018, pp. 1–10.

[170] Sven Mayer, Valentin Schwind, Robin Schweigert, and Niels Henze. “The Effect of Offset Correction and Cursor on Mid-Air Pointing in Real and Virtual Environments”. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI ’18*. the 2018 CHI Conference. Montreal QC, Canada: ACM Press, 2018, pp. 1–13.

[171] Victoria McArthur, Steven J. Castellucci, and I. Scott MacKenzie. “An empirical comparison of "wiimote" gun attachments for pointing tasks”. In: *Proceedings of the 1st ACM SIGCHI symposium on Engineering interactive computing systems*. EICS ’09. Pittsburgh, PA, USA: Association for Computing Machinery, July 15, 2009, pp. 203–208.

[172] David C. McCallum and Pourang Irani. “ARC-Pad: absolute+relative cursor positioning for large displays with a mobile touchscreen”. In: *Proceedings of the 22nd annual ACM symposium on User interface software and technology - UIST ’09*. the 22nd annual ACM symposium. Victoria, BC, Canada: ACM Press, 2009, p. 153.

[173] Peter Mohr, Markus Tatzgern, Jens Grubert, Dieter Schmalstieg, and Denis Kalkofen. “Adaptive user perspective rendering for Handheld Augmented Reality”. In: *2017 IEEE Symposium on 3D User Interfaces (3DUI)*. 2017 IEEE Symposium on 3D User Interfaces (3DUI). Los Angeles, CA, USA: IEEE, 2017, pp. 176–181.

[174] Peter Mohr, Markus Tatzgern, Tobias Langlotz, Andreas Lang, Dieter Schmalstieg, and Denis Kalkofen. “TrackCap: Enabling Smartphones for 3D Interaction on Mobile Head-Mounted Displays”. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI*

'19. the 2019 CHI Conference. Glasgow, Scotland Uk: ACM Press, 2019, pp. 1–11.

[175] M. Mohring, C. Lessig, and O. Bimber. “Video See-Through AR on Consumer Cell-Phones”. In: *Third IEEE and ACM International Symposium on Mixed and Augmented Reality*. Third IEEE and ACM International Symposium on Mixed and Augmented Reality. Arlington, VA, USA: IEEE, 2004, pp. 252–253.

[176] Annette Mossel, Benjamin Venditti, and Hannes Kaufmann. “3DTouch and HOMER-S: intuitive manipulation techniques for one-handed handheld augmented reality”. In: *Proceedings of the Virtual Reality International Conference on Laval Virtual - VRIC '13*. the Virtual Reality International Conference. Laval, France: ACM Press, 2013, p. 1.

[177] Brad A. Myers, Rishi Bhatnagar, Jeffrey Nichols, Choon Hong Peck, Dave Kong, Robert Miller, and A. Chris Long. “Interacting at a distance: measuring the performance of laser pointers and other devices”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '02. Minneapolis, Minnesota, USA: Association for Computing Machinery, Apr. 20, 2002, pp. 33–40.

[178] Brad A. Myers, Herb Stiel, and Robert Gargiulo. “Collaboration using multiple PDAs connected to a PC”. In: *Proceedings of the 1998 ACM conference on Computer supported cooperative work*. CSCW '98. Seattle, Washington, USA: Association for Computing Machinery, Nov. 1, 1998, pp. 285–294.

[179] Miguel A. Nacenta, Yemliha Kamber, Yizhou Qiang, and Per Ola Kristensson. “Memorability of Pre-Designed and User-Defined Gesture Sets”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '13. Paris, France: Association for Computing Machinery, 2013, pp. 1099–1108.

[180] Mathieu Nancel, Olivier Chapuis, Emmanuel Pietriga, Xing-Dong Yang, Pourang P. Irani, and Michel Beaudouin-Lafon. “High-precision pointing on large wall displays using small handheld devices”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*. the SIGCHI Conference. Paris, France: ACM Press, 2013, p. 831.

[181] Mathieu Nancel, Emmanuel Pietriga, Olivier Chapuis, and Michel Beaudouin-Lafon. “Mid-Air Pointing on Ultra-Walls”. In: *ACM Transactions on Computer-Human Interaction* 22.5 (Aug. 10, 2015), 21:1–21:62.

[182] Mathieu Nancel, Julie Wagner, Emmanuel Pietriga, Olivier Chapuis, and Wendy Mackay. “Mid-air pan-and-zoom on wall-sized displays”. In: *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI ’11*. the 2011 annual conference. Vancouver, BC, Canada: ACM Press, 2011, p. 177.

[183] Pai Chet Ng, James She, Kang Eun Jeon, and Matthias Baldauf. *When Smart Devices Interact With Pervasive Screens: A Survey*. Aug. 12, 2017. URL: https://doi.org/10.1145/3115933.

[184] Kai Nickel and Rainer Stiefelhagen. “Pointing gesture recognition based on 3D-tracking of face, hands and head orientation”. In: *Proceedings of the 5th international conference on Multimodal interfaces*. ICMI ’03. Vancouver, British Columbia, Canada: Association for Computing Machinery, Nov. 5, 2003, pp. 140–146.

[185] Donald A. Norman. “Natural User Interfaces Are Not Natural”. In: *Interactions* 17.3 (May 2010), pp. 6–10.

[186] Dan R. Olsen and Travis Nielsen. “Laser pointer interaction”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’01. Seattle, Washington, USA: Association for Computing Machinery, Mar. 1, 2001, pp. 17–22.

[187] Morin Ostkamp, Sven Heitmann, and Christian Kray. “Short-range optical interaction between smartphones and public displays”. In: *Proceedings of the 4th International Symposium on Pervasive Displays - PerDis ’15*. the 4th International Symposium. Saarbruecken, Germany: ACM Press, 2015, pp. 39–46.

[188] Jeni Paay, Jesper Kjeldskov, Dimitrios Raptis, Mikael B. Skov, Ivan S. Penchev, and Elias Ringhauge. “Cross-device interaction with large displays in public: insights from both users’ and observers’ perspectives”. In: *Proceedings of the 29th Australian Conference on Computer-Human Interaction - OZCHI ’17*. the 29th Australian Conference. Brisbane, Queensland, Australia: ACM Press, 2017, pp. 87–97.

[189] Jeni Paay, Dimitrios Raptis, Jesper Kjeldskov, Mikael B. Skov, Eric V. Ruder, and Bjarke M. Lauridsen. “Investigating Cross-Device Interaction between a Handheld Device and a Large Display”. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI ’17*. the 2017 CHI Conference. Denver, Colorado, USA: ACM Press, 2017, pp. 6608–6619.

[190] Michel Pahud, Ken Hinckley, Shamsi Iqbal, Abigail Sellen, and Bill Buxton. “Toward compound navigation tasks on mobiles via spatial manipulation”. In: *Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services - MobileHCI ’13*. the 15th international conference. Munich, Germany: ACM Press, 2013, p. 113.

[191] Patrick Parzer, Kathrin Probst, Teo Babic, Christian Rendl, Anita Vogl, Alex Olwal, and Michael Haller. “FlexTiles: A Flexible, Stretchable, Formable, Pressure-Sensitive, Tactile Input Sensor”. 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, May 7, 2016, pp. 3754–3757.

[192] Patchara Pattanakimhun, Winyu Chinthammit, and Nopporn Chotikakamthorn. “Enhanced engagement with public displays through mobile phone interaction”. In: *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications*. SA ’17. Bangkok, Thailand: Association for Computing Machinery, Nov. 27, 2017, pp. 1–5.

[193] Patchara Pattanakimhun, Winyu Chinthammit, and Nopporn Chotikakamthorn. “Evaluation of mobile phone interaction with large public displays”. In: *2017 14th International Joint Conference on Computer Science and Software Engineering (JCSSE)*. 2017 14th International Joint Conference on Computer Science and Software Engineering (JCSSE). NakhonSiThammarat, Thailand: IEEE, July 2017, pp. 1–6.

[194] Choon Hong Peck. “Useful parameters for the design of laser pointer interaction techniques”. In: *CHI ’01 Extended Abstracts on Human Factors in Computing Systems*. CHI EA ’01. Seattle, Washington: Association for Computing Machinery, Mar. 31, 2001, pp. 461–462.

[195] Bruno Perez. *An accessible guide to inclusive design*. URL: https://medium.com/@brunoperez/an-accessible-guide-to-inclusive-design-91b36a227819.

[196] Julian Petford, Miguel A. Nacenta, and Carl Gutwin. “Pointing All Around You: Selection Performance of Mouse and Ray-Cast Pointing in Full-Coverage Displays”. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI ’18. Montreal QC, Canada: Association for Computing Machinery, Apr. 21, 2018, pp. 1–14.

[197] Ken Pfeuffer and Hans Gellersen. “Gaze and Touch Interaction on Tablets”. In: *Proceedings of the 29th Annual Symposium on User Interface Software and Technology - UIST ’16*. the 29th Annual Symposium. Tokyo, Japan: ACM Press, 2016, pp. 301–311.

[198] Ken Pfeuffer, Ken Hinckley, Michel Pahud, and Bill Buxton. “Thumb + Pen Interaction on Tablets”. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI ’17*. the 2017 CHI Conference. Denver, Colorado, USA: ACM Press, 2017, pp. 3254–3266.

[199] Ken Pfeuffer, Melodie Vidal, Jayson Turner, Andreas Bulling, and Hans Gellersen. “Pursuit calibration: making gaze calibration less tedious and more flexible”. In: *Proceedings of the 26th annual ACM symposium on User interface software and technology - UIST ’13*. the 26th annual ACM symposium. St. Andrews, Scotland, United Kingdom: ACM Press, 2013, pp. 261–270.

[200] Krzysztof Pietroszek, Anastasia Kuzminykh, James R. Wallace, and Edward Lank. “Smartcasting: a discount 3D interaction technique for public displays”. In: *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures the Future of Design - OzCHI ’14*. the 26th Australian Computer-Human Interaction Conference. Sydney, New South Wales, Australia: ACM Press, 2014, pp. 119–128.

[201] Krzysztof Pietroszek, James R. Wallace, and Edward Lank. “Tiltcasting: 3D Interaction on Large Displays using a Mobile Device”. In: *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology - UIST ’15*. the 28th Annual ACM Symposium. Daegu, Kyungpook, Republic of Korea: ACM Press, 2015, pp. 57–62.

[202] Alexandros Pino, Evangelos Tzemis, Nikolaos Ioannou, and Georgios Kouroupetroglou. “Using Kinect for 2D and 3D Pointing Tasks: Performance Evaluation”. In: *Proceedings of the 15th International Conference on Human-Computer Interaction: Interaction Modalities and Techniques -*

Volume Part IV. HCI'13. Las Vegas, NV: Springer-Verlag, 2013, pp. 358–367.

[203] Arnaud Prouzeau, Maxime Cordeil, Clement Robin, Barrett Ens, Bruce H. Thomas, and Tim Dwyer. “Scaptics and Highlight-Planes: Immersive Interaction Techniques for Finding Occluded Features in 3D Scatterplots”. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*. the 2019 CHI Conference. Glasgow, Scotland Uk: ACM Press, 2019, pp. 1–12.

[204] Jing Qian, Angel Cheung, Meredith Young-Ng, Fumeng Yang, Xiangyu Li, and Jeff Huang. “Portalware: A Smartphone-Wearable Dual-Display System for Expanding the Free-Hand Interaction Region in Augmented Reality”. In: (2020), p. 8.

[205] Roman Rädle, Hans-Christian Jetter, Simon Butscher, and Harald Reiterer. “The effect of egocentric body movements on users' navigation performance and spatial memory in zoomable user interfaces”. In: *Proceedings of the 2013 ACM international conference on Interactive tabletops and surfaces - ITS '13*. the 2013 ACM international conference. St. Andrews, Scotland, United Kingdom: ACM Press, 2013, pp. 23–32.

[206] Roman Rädle, Hans-Christian Jetter, Nicolai Marquardt, Harald Reiterer, and Yvonne Rogers. “HuddleLamp: Spatially-Aware Mobile Displays for Ad-hoc Around-the-Table Collaboration”. In: *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces*. ITS '14. Dresden, Germany: Association for Computing Machinery, Nov. 16, 2014, pp. 45–54.

[207] Roman Rädle, Hans-Christian Jetter, Mario Schreiner, Zhihao Lu, Harald Reiterer, and Yvonne Rogers. “Spatially-aware or Spatially-agnostic? Elicitation and Evaluation of User-Defined Cross-Device Interactions”. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI '15. Seoul, Republic of Korea: Association for Computing Machinery, Apr. 18, 2015, pp. 3913–3922.

[208] Mathieu Raynal and Guillaume Gauffre. “Tactile camera vs. tangible camera: taking advantage of small physical artefacts to navigate into large data collection”. In: (2010), p. 10.

[209] J. Rekimoto. “A vision-based head tracker for fish tank virtual reality-VR without head gear”. In: *Proceedings Virtual Reality Annual International Symposium ’95*. Virtual Reality Annual International Symposium ’95. Research Triangle Park, NC, USA: IEEE Comput. Soc. Press, 1995, pp. 94–100.

[210] J. Rekimoto. “Matrix: a realtime object identification and registration method for augmented reality”. In: *Proceedings. 3rd Asia Pacific Computer Human Interaction (Cat. No.98EX110)*. 1998, pp. 63–68.

[211] Jun Rekimoto. “Pick-and-drop: a direct manipulation technique for multiple computer environments”. In: *Proceedings of the 10th annual ACM symposium on User interface software and technology*. UIST ’97. Banff, Alberta, Canada: Association for Computing Machinery, Oct. 1, 1997, pp. 31–39.

[212] Jun Rekimoto. “Tilting Operations for Small Screen Interfaces (Tech Note)”. In: (1996), p. 2.

[213] Radiah Rivu, Mariam Hassib, Yasmeen Abdrabou, Florian Alt, and Ken Pfeuffer. “Gaze’N’Touch: Enhancing Text Selection on Mobile Devices Using Gaze”. In: (2020), p. 8.

[214] Michael Rohs. “Linking Physical and Virtual Worlds with Visual Markers and Handheld Devices”. In: (2005), p. 236.

[215] Michael Rohs and Antti Oulasvirta. “Target acquisition with camera phones when used as magic lenses”. In: *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI ’08*. Proceeding of the twenty-sixth annual CHI conference. Florence, Italy: ACM Press, 2008, p. 1409.

[216] Jaime Ruiz, Yang Li, and Edward Lank. “User-defined motion gestures for mobile interaction”. In: *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI ’11*. the 2011 annual conference. Vancouver, BC, Canada: ACM Press, 2011, p. 197.

[217] Jaime Ruiz and Daniel Vogel. “Soft-Constraints to Reduce Legacy and Performance Bias to Elicit Whole-body Gestures with Low Arm Fatigue”. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI ’15*. the 33rd Annual ACM Conference. Seoul, Republic of Korea: ACM Press, 2015, pp. 3347–3350.

[218] MacKenzie I. S. “The Wiley Handbook of Human Computer Interaction (eds K. L. Norman and J. Kirakowski)”. In: Chichester, UK: John Wiley & Sons, Ltd, 2018. Chap. Fitts’ Law.

[219] R. F. Salas-Moreno, B. Glocken, P. H. J. Kelly, and A. J. Davison. “Dense planar SLAM”. In: *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 2014, pp. 157–164.

[220] Thiago Santini, Wolfgang Fuhl, and Enkelejda Kasneci. “CalibMe: Fast and Unsupervised Eye Tracker Calibration for Gaze-Based Pervasive Human-Computer Interaction”. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. CHI ’17: CHI Conference on Human Factors in Computing Systems. Denver Colorado USA: ACM, May 2, 2017, pp. 2594–2605.

[221] T. Scott Saponas, Chris Harrison, and Hrvoje Benko. “PocketTouch: through-fabric capacitive touch input”. In: *Proceedings of the 24th annual ACM symposium on User interface software and technology - UIST ’11*. the 24th annual ACM symposium. Santa Barbara, California, USA: ACM Press, 2011, p. 303.

[222] Jürgen Scheible, Timo Ojala, and Paul Coulton. “MobiToss: a novel gesture based interface for creating and sharing mobile multimedia art on large public displays”. In: *Proceedings of the 16th ACM international conference on Multimedia*. MM ’08. Vancouver, British Columbia, Canada: Association for Computing Machinery, Oct. 26, 2008, pp. 957–960.

[223] D. Schmalstieg and Tobias Höllerer. *Augmented reality: principles and practice*. Addison-Wesley usability and HCI series. OCLC: ocn802326915. Boston: Addison-Wesley, 2016. 496 pp.

[224] Philip Schuchardt and Doug A. Bowman. “The benefits of immersion for spatial understanding of complex underground cave systems”. In: *Proceedings of the 2007 ACM symposium on Virtual reality software and technology - VRST ’07*. the 2007 ACM symposium. Newport Beach, California: ACM Press, 2007, p. 121.

[225] Adriano Scoditti, Renaud Blanch, and Joëlle Coutaz. “A novel taxonomy for gestural interaction techniques based on accelerometers”. In: *Proceedings of the 15th international conference on Intelligent user interfaces - IUI ’11*.

the 15th international conference. Palo Alto, CA, USA: ACM Press, 2011, p. 63.

[226] Julian Seifert, Andreas Bayer, and Enrico Rukzio. “PointerPhone: Using Mobile Phones for Direct Pointing Interactions with Remote Displays”. In: *Human-Computer Interaction – INTERACT 2013*. Ed. by Paula Kotzé, Gary Marsden, Gitte Lindgaard, Janet Wesson, and Marco Winckler. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2013, pp. 18–35.

[227] Marcos Serrano, Eric Lecolinet, and Yves Guiard. “Bezel-Tap gestures: quick activation of commands from sleep mode on tablets”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI ’13*. the SIGCHI Conference. Paris, France: ACM Press, 2013, p. 3027.

[228] Teddy Seyed, Xing-Dong Yang, and Daniel Vogel. “A Modular Smartphone for Lending”. In: *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology - UIST ’17*. the 30th Annual ACM Symposium. Quebec City, QC, Canada: ACM Press, 2017, pp. 205–215.

[229] Garth Shoemaker, Anthony Tang, and Kellogg S Booth. “Shadow Reaching: A New Perspective on Interaction for Large Wall Displays”. In: (2007), p. 4.

[230] Garth Shoemaker, Takayuki Tsukitani, Yoshifumi Kitamura, and Kellogg S. Booth. “Body-centric interaction techniques for very large wall displays”. In: *Proceedings of the 6th Nordic Conference on Human-Computer Interaction Extending Boundaries - NordiCHI ’10*. the 6th Nordic Conference. Reykjavik, Iceland: ACM Press, 2010, p. 463.

[231] Linda E. Sibert and Robert J. K. Jacob. “Evaluation of eye gaze interaction”. In: *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI ’00*. the SIGCHI conference. The Hague, The Netherlands: ACM Press, 2000, pp. 281–288.

[232] Bruno Siciliano. *Springer handbook of robotics*. Berlin: Springer, 2008.

[233] Shaishav Siddhpuria, Keiko Katsuragawa, James R. Wallace, and Edward Lank. “Exploring At-Your-Side Gestural Interaction for Ubiquitous Environments”. In: *Proceedings of the 2017 Conference on Designing Interactive Systems - DIS ’17*. the 2017 Conference. Edinburgh, United Kingdom: ACM Press, 2017, pp. 1111–1122.

[234] Shaishav Siddhpuria, Sylvain Malacria, Mathieu Nancel, and Edward Lank. “Pointing at a Distance with Everyday Smart Devices”. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI ’18. Montreal QC, Canada: Association for Computing Machinery, Apr. 19, 2018, pp. 1–11.

[235] ACM SIGCHI. *Youtube. CHI 2017 Tuesday keynote - Ben Shneiderman. Video Timestamp: 50min 53sec*. URL: https://www.youtube.com/watch?v=eI1yCZXvS5s&t=3053s.

[236] Miika Silfverberg, I Scott MacKenzie, and Tatu Kauppinen. “An Isometric Joystick as a Pointing Device for Handheld Information Terminals”. In: (2001), p. 8.

[237] Adalberto L. Simeone, Julian Seifert, Dominik Schmidt, Paul Holleis, Enrico Rukzio, and Hans Gellersen. “A cross-device drag-and-drop technique”. In: *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia - MUM ’13*. the 12th International Conference. Luleå, Sweden: ACM Press, 2013, pp. 1–4.

[238] Brian A. Smith, Qi Yin, Steven K. Feiner, and Shree K. Nayar. “Gaze locking: passive eye contact detection for human-object interaction”. In: *Proceedings of the 26th annual ACM symposium on User interface software and technology - UIST ’13*. the 26th annual ACM symposium. St. Andrews, Scotland, United Kingdom: ACM Press, 2013, pp. 271–280.

[239] Washington State Department of Social and Range of Joint Motion Evaluation Chart Health Services. URL: https://www.etas.ee/wp-content/uploads/2013/10/ROM-normid.pdf.

[240] Rajinder Sodhi, Hrvoje Benko, and Andrew Wilson. “LightGuide: projected visualizations for hand movement guidance”. In: *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI ’12*. the 2012 ACM annual conference. Austin, Texas, USA: ACM Press, 2012, p. 179.

[241] Hendrik Sollich, Ulrich von Zadow, Tobias Pietzsch, Pavel Tomancak, and Raimund Dachselt. “Exploring Time-dependent Scientific Data Using Spatially Aware Mobiles and Large Displays”. In: *Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces*. ISS

’16. Niagara Falls, Ontario, Canada: Association for Computing Machinery, Nov. 6, 2016, pp. 349–354.

[242] Peng Song, Wooi Boon Goh, Chi-Wing Fu, Qiang Meng, and Pheng-Ann Heng. “WYSIWYF: exploring and annotating volume data with a tangible handheld device”. In: *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI ’11*. the 2011 annual conference. Vancouver, BC, Canada: ACM Press, 2011, p. 1333.

[243] Oleg Špakov, Poika Isokoski, and Päivi Majaranta. “Look and lean: accurate head-assisted eye pointing”. In: *Proceedings of the Symposium on Eye Tracking Research and Applications - ETRA ’14*. the Symposium. Safety Harbor, Florida: ACM Press, 2014, pp. 35–42.

[244] Marco Speicher, Florian Daiber, Sven Gehring, and Antonio Krüger. “Exploring 3D manipulation on large stereoscopic displays”. In: *Proceedings of the 5th ACM International Symposium on Pervasive Displays - PerDis ’16*. the 5th ACM International Symposium. Oulu, Finland: ACM Press, 2016, pp. 59–66.

[245] Martin Spindler, Wolfgang Büschel, and Raimund Dachselt. “Use your head: tangible windows for 3D information spaces in a tabletop environment”. In: *Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces - ITS ’12*. the 2012 ACM international conference. Cambridge, Massachusetts, USA: ACM Press, 2012, p. 245.

[246] Martin Spindler, Martin Schuessler, Marcel Martsch, and Raimund Dachselt. “Pinch-drag-flick vs. spatial input: rethinking zoom & pan on mobile displays”. In: *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI ’14*. the 32nd annual ACM conference. Toronto, Ontario, Canada: ACM Press, 2014, pp. 1113–1122.

[247] Statista. *Global market share held by mobile operating systems from 2009 to 2023*. URL: https://www.statista.com/statistics/272698/global-market-share-held-by-mobile-operating-systems-since-2009/.

[248] Sophie Stellmach and Raimund Dachselt. “Look & touch: gaze-supported target acquisition”. In: *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI ’12*. the 2012 ACM annual conference. Austin, Texas, USA: ACM Press, 2012, p. 2981.

[249] Sophie Stellmach and Raimund Dachselt. “Looking at 3D User Interfaces”. In: (2012), p. 4.

[250] Sophie Stellmach and Raimund Dachselt. “Still looking: investigating seamless gaze-supported selection, positioning, and manipulation of distant targets”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’13. Paris, France: Association for Computing Machinery, Apr. 27, 2013, pp. 285–294.

[251] Constantine Stephanidis and Demosthenes Akoumianakis. “Universal Design: Towards Universal Access in the Information Society”. In: *CHI ’01 Extended Abstracts on Human Factors in Computing Systems*. CHI EA ’01. Seattle, Washington: Association for Computing Machinery, 2001, pp. 499–500.

[252] Constantine Stephanidis, Demosthenes Akoumianakis, and Anthony Savidis. “Universal Design in Human-Computer Interaction”. In: Jan. 2001, pp. 741–745.

[253] Molly Follette Story. “Maximizing Usability: The Principles of Universal Design”. In: *Assistive Technology* 10.1 (1998). PMID: 10181150, pp. 4–12. eprint: https://doi.org/10.1080/10400435.1998.10131955.

[254] Yusuke Sugano, Xucong Zhang, and Andreas Bulling. “AggreGaze: Collective Estimation of Audience Attention on Public Displays”. In: *Proceedings of the 29th Annual Symposium on User Interface Software and Technology - UIST ’16*. the 29th Annual Symposium. Tokyo, Japan: ACM Press, 2016, pp. 821–831.

[255] Hemant Bhaskar Surale, Aakar Gupta, Mark Hancock, and Daniel Vogel. “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*. the 2019 CHI Conference. Glasgow, Scotland Uk: ACM Press, 2019, pp. 1–13.

[256] Robert J Teather and I Scott MacKenzie. “Position vs. Velocity Control for Tilt-Based Interaction”. In: (2014), p. 8.

[257] Robert J. Teather and Wolfgang Stuerzlinger. “Pointing at 3d target projections with one-eyed and stereo cursors”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI ’13*. the SIGCHI Conference. Paris, France: ACM Press, 2013, p. 159.

[258] Hsin-Ruey Tsai, Te-Yen Wu, Da-Yuan Huang, Min-Chieh Hsiu, Jui-Chun Hsiao, Yi-Ping Hung, Mike Y. Chen, and Bing-Yu Chen. “SegTouch: Enhancing Touch Input While Providing Touch Gestures on Screens Using Thumb-To-Index-Finger Gestures”. In: *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA ’17*. the 2017 CHI Conference Extended Abstracts. Denver, Colorado, USA: ACM Press, 2017, pp. 2164–2171.

[259] Jayson Turner, Jason Alexander, Andreas Bulling, and Hans Gellersen. “Gaze+RST: Integrating Gaze and Multitouch for Remote Rotate-Scale-Translate Tasks”. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI ’15*. the 33rd Annual ACM Conference. Seoul, Republic of Korea: ACM Press, 2015, pp. 4179–4188.

[260] Jayson Turner, Andreas Bulling, Jason Alexander, and Hans Gellersen. “Eye drop: an interaction concept for gaze-supported point-to-point content transfer”. In: *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia - MUM ’13*. the 12th International Conference. Luleå, Sweden: ACM Press, 2013, pp. 1–4.

[261] B. Ullmer and H. Ishii. “Emerging frameworks for tangible user interfaces”. In: *IBM Systems Journal* 39.3 (2000), pp. 915–931.

[262] Jouni Vepsäläinen, Antonella Di Rienzo, Matti Nelimarkka, Jouni A. Ojala, Petri Savolainen, Kai Kuikkaniemi, Sasu Tarkoma, and Giulio Jacucci. “Personal Device as a Controller for Interactive Surfaces: Usability and Utility of Different Connection Methods”. In: *Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces*. ITS ’15. Madeira, Portugal: Association for Computing Machinery, Nov. 15, 2015, pp. 201–204.

[263] Vinayak, Devarajan Ramanujan, Cecil Piya, and Karthik Ramani. “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*. the TEI ’16: Tenth International Conference. Eindhoven, Netherlands: ACM Press, 2016, pp. 12–20.

[264] Thomas Vincent, Laurence Nigay, and Takeshi Kurata. “Handheld Augmented Reality: Effect of registration jitter on cursor-based pointing techniques”. In: *Proceedings of the 25th ICME conference francophone on*

l'Interaction Homme-Machine - IHM '13. the 25th IEME conference francophone. Talence, France: ACM Press, 2013, pp. 1–6.

[265] HTC Vive. *Vive Tracker*. URL: https://www.vive.com/eu/vive-tracker/.

[266] Simon Voelker, Sebastian Hueber, Christian Corsten, and Christian Remy. “HeadReach: Using Head Tracking to Increase Reachability on Mobile Touch Devices”. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI '20. Honolulu, HI, USA: Association for Computing Machinery, Apr. 21, 2020, pp. 1–12.

[267] Simon Voelker, Sebastian Hueber, Christian Holz, Christian Remy, and Nicolai Marquardt. “GazeConduits: Calibration-Free Cross-Device Collaboration through Gaze and Touch”. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI '20. Honolulu, HI, USA: Association for Computing Machinery, Apr. 21, 2020, pp. 1–10.

[268] Daniel Vogel and Ravin Balakrishnan. “Distant freehand pointing and clicking on very large, high resolution displays”. In: *Proceedings of the 18th annual ACM symposium on User interface software and technology - UIST '05*. the 18th annual ACM symposium. Seattle, WA, USA: ACM Press, 2005, p. 33.

[269] Daniel Vogel and Ravin Balakrishnan. “Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users”. In: *Proceedings of the 17th annual ACM symposium on User interface software and technology*. UIST '04. Santa Fe, NM, USA: Association for Computing Machinery, Oct. 24, 2004, pp. 137–146.

[270] F. Vogt, J. Wong, B.A. Po, R. Argue, S.S. Fels, and K.S. Booth. “Exploring collaboration with group pointer interaction”. In: *Proceedings Computer Graphics International, 2004.* Proceedings Computer Graphics International, 2004. Crete, Greece: IEEE, 2004, pp. 636–639.

[271] Vanessa Vuibert, Wolfgang Stuerzlinger, and Jeremy R. Cooperstock. “Evaluation of Docking Task Performance Using Mid-air Interaction Techniques”. In: *Proceedings of the 3rd ACM Symposium on Spatial User Interaction*. SUI '15. Los Angeles, California, USA: ACM, 2015, pp. 44–52.

[272] Philipp Wacker, Oliver Nowak, Simon Voelker, and Jan Borchers. “ARPen: Mid-Air Object Manipulation Techniques for a Bimanual AR System with Pen & Smartphone”. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI ’19*. the 2019 CHI Conference. Glasgow, Scotland Uk: ACM Press, 2019, pp. 1–12.

[273] Philipp Wacker, Adrian Wagner, Simon Voelker, and Jan Borchers. “Heatmaps, Shadows, Bubbles, Rays: Comparing Mid-Air Pen Position Visualizations in Handheld AR”. In: (2020), p. 11.

[274] D. Wagner and D. Schmalstieg. “First steps towards handheld augmented reality”. In: *Seventh IEEE International Symposium on Wearable Computers, 2003. Proceedings.* Seventh IEEE International Symposium on Wearable Computers, 2003. White Plains, NY, USA: IEEE, 2003, pp. 127–135.

[275] Daniel Wagner, Tobias Langlotz, and Dieter Schmalstieg. “Robust and unobtrusive marker tracking on mobile phones”. In: *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*. 2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR). Cambridge, UK: IEEE, Sept. 2008, pp. 121–124.

[276] Daniel Wagner and Dieter Schmalstieg. “History and Future of Tracking for Mobile Phone Augmented Reality”. In: *2009 International Symposium on Ubiquitous Virtual Reality*. 2009 International Symposium on Ubiquitous Virtual Reality (ISUVR). GIST, Guangju, Korea: IEEE, July 2009, pp. 7–10.

[277] Daniel Wagner, Dieter Schmalstieg, and Horst Bischof. “Multiple target detection and tracking with guaranteed framerates on mobile phones”. In: *2009 8th IEEE International Symposium on Mixed and Augmented Reality*. 2009 8th IEEE International Symposium on Mixed and Augmented Reality (ISMAR). Orlando, FL, USA: IEEE, Oct. 2009, pp. 57–64.

[278] Robert Walter, Gilles Bailly, and Jörg Müller. “StrikeAPose: revealing mid-air gestures on public displays”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI ’13*. the SIGCHI Conference. Paris, France: ACM Press, 2013, p. 841.

[279] Robert Walter, Gilles Bailly, Nina Valkanova, and Jörg Müller. “Cuenesics: using mid-air gestures to select items on interactive public displays”. In: *Proceedings of the 16th international conference on Human-computer*

interaction with mobile devices & services. MobileHCI ’14. Toronto, ON, Canada: Association for Computing Machinery, Sept. 23, 2014, pp. 299–308.

[280] Jingtao Wang, Shumin Zhai, and John Canny. “Camera Phone Based Motion Sensing: Interaction Techniques, Applications and Performance Study”. In: (2006), p. 10.

[281] C. Ware and D. Fleet. “Integrating flying and fish tank metaphors with cyclopean scale”. In: *Proceedings Computer Graphics International*. Computer Graphics International. Hasselt and Diepenbeek, Belgium: IEEE Comput. Soc. Press, 1997, pp. 39–46.

[282] Colin Ware, Kevin Arthur, and Kellogg S. Booth. “Fish tank virtual reality”. In: *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI ’93*. the SIGCHI conference. Amsterdam, The Netherlands: ACM Press, 1993, pp. 37–42.

[283] Rina R Wehbe, Terence Dickson, Anastasia Kuzminykh, Lennart E Nacke, and Edward Lank. “Personal Space in Play: Physical and Digital Boundaries in Large-Display Cooperative and Competitive Games”. In: (2020), p. 14.

[284] Daniel Wigdor and Dennis Wixon. *Brave NUI World: Designing Natural User Interfaces for Touch and Gesture*. 1st. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2011.

[285] Karl D.D. Willis, Ivan Poupyrev, and Takaaki Shiratori. “Motionbeam: a metaphor for character interaction with handheld projectors”. In: *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI ’11*. the 2011 annual conference. Vancouver, BC, Canada: ACM Press, 2011, p. 1031.

[286] Andrew Wilson and Steven Shafer. “XWand: UI for Intelligent Spaces”. In: *NEW HORIZONS* 5 (2003), p. 8.

[287] Christian Winkler, Ken Pfeuffer, and Enrico Rukzio. “Investigating mid-air pointing interaction for projector phones”. In: *Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces*. ITS ’12. Cambridge, Massachusetts, USA: Association for Computing Machinery, Nov. 11, 2012, pp. 85–94.

[288] Jiahui Wu, Gang Pan, Daqing Zhang, Guande Qi, and Shijian Li. “Gesture Recognition with a 3-D Accelerometer”. In: (2009), p. 14.

[289] Wataru Yamada, Hiroyuki Manabe, and Daizo Ikeda. “CamTrackPoint: Camera-Based Pointing Stick Using Transmitted Light through Finger”. In: *The 31st Annual ACM Symposium on User Interface Software and Technology - UIST ’18*. The 31st Annual ACM Symposium. Berlin, Germany: ACM Press, 2018, pp. 313–320.

[290] Ka-Ping Yee. “Peephole Displays: Pen Interaction on Spatially Aware Handheld Computers”. In: (2003), p. 8.

[291] Hui-Shyong Yeo, Juyoung Lee, Hyung-il Kim, Aakar Gupta, Andrea Bianchi, Daniel Vogel, Hideki Koike, Woontack Woo, and Aaron Quigley. “WRIST: Watch-Ring Interaction and Sensing Technique for Wrist Gestures and Macro-Micro Pointing”. In: *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services*. MobileHCI ’19. Taipei, Taiwan: Association for Computing Machinery, Oct. 1, 2019, pp. 1–15.

[292] Ulrich von Zadow. “Using Personal Devices to Facilitate Multi-user Interaction with Large Display Walls”. In: *Adjunct Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. UIST ’15 Adjunct. Daegu, Kyungpook, Republic of Korea: Association for Computing Machinery, Nov. 6, 2015, pp. 25–28.

[293] Ulrich von Zadow, Wolfgang Büschel, Ricardo Langner, and Raimund Interactive Media Lab Dachselt. “SleeD: Using a Sleeve Display to Interact with Touch-sensitive Display Walls”. In: *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces - ITS ’14*. the Ninth ACM International Conference. Dresden, Germany: ACM Press, 2014, pp. 129–138.

[294] Ulrich von Zadow, Patrick Reipschläger, Daniel Bösel, Anita Sellent, and Raimund Dachselt. “YouTouch! Low-Cost User Identification at an Interactive Display Wall”. In: *Proceedings of the International Working Conference on Advanced Visual Interfaces*. AVI ’16. Bari, Italy: Association for Computing Machinery, June 7, 2016, pp. 144–151.

[295] Ulrich von Zadow, Andreas Siegel, and Raimund Dachselt. “Multi-user Multi-device Interaction with Large Displays at the Point of Sale: An Application Case”. In: *Proceedings of the 2015 International Conference on*

Interactive Tabletops & Surfaces - ITS '15. the 2015 International Conference. Madeira, Portugal: ACM Press, 2015, pp. 343–348.

[296] Shumin Zhai, William Buxton, and Paul Milgram. “The partial-occlusion effect: utilizing semitransparency in 3D human-computer interaction”. In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 3.3 (Sept. 1996), pp. 254–284.

[297] Shumin Zhai and Paul Milgram. “Quantifying coordination in multiple DOF movement and its application to evaluating 6 DOF input devices”. In: *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '98*. the SIGCHI conference. Los Angeles, California, United States: ACM Press, 1998, pp. 320–327.

[298] Shumin Zhai, Carlos Morimoto, and Steven Ihde. “Manual and gaze input cascaded (MAGIC) pointing”. In: *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. CHI '99. Pittsburgh, Pennsylvania, USA: Association for Computing Machinery, May 1, 1999, pp. 246–253.

[299] Boning Zhang, Yiqiang Chen, Yueliang Qian, and Xiangdong Wang. “A ring-shaped interactive device for large remote display and mobile device control”. In: *Proceedings of the 13th international conference on Ubiquitous computing*. UbiComp '11. Beijing, China: Association for Computing Machinery, Sept. 17, 2011, pp. 473–474.

[300] Xiaoyi Zhang, Harish Kulkarni, and Meredith Ringel Morris. “Smartphone-Based Gaze Gesture Communication for People with Motor Disabilities”. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. CHI '17: CHI Conference on Human Factors in Computing Systems. Denver Colorado USA: ACM, May 2, 2017, pp. 2878–2889.

[301] Xucong Zhang, Yusuke Sugano, and Andreas Bulling. “Evaluation of Appearance-Based Methods and Implications for Gaze-Based Applications”. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*. the 2019 CHI Conference. Glasgow, Scotland Uk: ACM Press, 2019, pp. 1–13.

[302] Yu Zhong, Xin Li, Mingming Fan, and Yuanchun Shi. “Doodle space: painting on a public display by cam-phone”. In: *Proceedings of the 2009 workshop on Ambient media computing - AMC '09*. the 2009 workshop. Beijing, China: ACM Press, 2009, p. 13.

[303] Fengyuan Zhu and Tovi Grossman. “BISHARE: Exploring Bidirectional Interactions Between Smartphones and Head-Mounted Augmented Reality”. In: (2020), p. 14.

[304] Jan Zizka, Alex Olwal, and Ramesh Raskar. “SpeckleSense: fast, precise, low-cost and compact motion sensing using laser speckle”. In: *Proceedings of the 24th annual ACM symposium on User interface software and technology - UIST ’11*. the 24th annual ACM symposium. Santa Barbara, California, USA: ACM Press, 2011, p. 489.