Université de Lille & University of Waterloo

Thèse

pour obtenir le grade de :

Docteur de l'Université de Lille & University of Waterloo

en informatique, spécialité en

« Interaction Humain-Machine »

par

Yuan Chen

Informatique de bureau pervasive par manipulation directe d'une lampe augmentée

Thèse soutenue le 20 Septembre 2024 devant le jury composé de :

President *Président*	Mark Hancock	Professor, Dept. of Management Sciences, University of Waterloo
External examiners *Rapporteurs*	Scott Bateman	Associate Professor, Faculty of Computer Science, University of New Brunswick
	Marcos Serrano	Associate Professor, University of Toulouse 3
Supervisors *Directeurs*	Géry Casiez	Professor, Université de Lille
	Edward Lank	Professor, Cheriton School of Computer Science, University of Waterloo
	Sylvain Malacria	Research Scientist, Inria centre at the University of Lille
	Daniel Vogel	Professor, Cheriton School of Computer Science, University of Waterloo
External members *Examinateurs*	Caroline Appert	Research scientist (DR CNRS), Université Paris Saclay
	Nadine Couture	Professor, Ecole Supérieure des Technologies Industrielles Avancées

AUTHOR'S DECLARATION

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

STATEMENT OF CONTRIBUTIONS

This dissertation includes first-authored peer-reviewed material that has appeared in conference and journal proceedings published by the Association for Computing Machinery (ACM). The ACM's policy is as follows[1]:

> *"Authors can include partial or complete papers of their own (and no fee is expected) in a dissertation as long as citations and DOI pointers to the Versions of Record in the ACM Digital Library are included. Authors can use any portion of their own work in presentations and in the classroom (and no fee is expected)."*

The following list serves as a declaration of the works included in this dissertation. This material is expanded and revised from the original publication.

Portions of Chapter 3
Yuan Chen, Géry Casiez, Sylvain Malacria, and Daniel Vogel. 2024. LuxAR: A Direct Manipulation Projected Display to Extend and Augment Desktop Computing. In Graphics Interface (GI '24). Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3670947.3670981

Portions of Chapter 5
Yuan Chen, Géry Casiez, Sylvain Malacria, and Edward Lank. 2023. Exploring the Effects of Intended Use on Targeting in Virtual Reality. In Graphics Interface (GI '23). 1-11. Canadian Information Processing Society.

1 https://authors.acm.org/author-resources/author-rights

ABSTRACT

Desktop computing, despite its long-standing dominance in personal productivity, remains largely confined to screens. Many efforts to expand beyond a single screen, from multiple monitors to incorporating projector-camera units or head-mounted displays, have shown promise. However, this is often from the desktop display to other devices and it lacks the awareness of physical environments and user activities. This thesis explores a novel form of direct manipulation projector-camera system, which leverages unique characteristics of physical lamp movement to manipulate content to and from the desktop display, but also to and from devices and the physical environment, while maintaining the awareness in the workspace.

Three projects examine the design, prototyping, and human factors aspects of an augmented lamp system in which the lamp works as an input and output device connecting desktop computing and physical environment. In the first project, an interaction design space is introduced for physical direct manipulation using an architect lamp with a proof-of-concept system using a projector and motion tracking system. We demonstrate its potential usage through three scenarios, describe study results evaluating its potential, and summarize design implications. In the second project, we study the impact on user performance and interaction strategies when interacting with an augmented lamp in a desktop space. We conduct a controlled experiment in Virtual Reality to examine two control mechanisms for target acquisition tasks in a dynamic peephole display: "coupled", when the display centre is used for selection, and "decoupled", when the selection is handled by separate inputs like direct touch. We find that the two control mechanisms have subtle differences in total time and error, but with the same technique, people show different kinematics patterns for coordinating the movement of a dynamic peephole display for searching the target. In the third project, we explore the latter observation in a more general context. Using a controlled Virtual Reality environment, we conduct an experiment to investigate whether what users intend to do with a virtual target impacts how they plan and perform the initial target acquisition. Our results lead to an understanding of user motion profiles before acquisition for different intended interactions with the same target. We discuss how these kinematics profiles can then be used to improve the lamp design, such as integrating force sensors into the lamp to improve its activity awareness. Together, these findings establish a promising way to connect current desktop computing with the surrounding physical desktop environment based on a deeper understanding of user activities in that space.

L'informatique de bureau, qui reste le moyen principal de réaliser efficacement une large gamme de tâches, continue d'afficher des informations confinées aux écrans. De nombreuses tentatives d'affichage au-delà d'un seul écran se sont révélées prometteuses, allant de l'utilisation de plusieurs écrans à l'utilisation de projecteurs ou de casques de réalité mixte. Cependant, cette extension se fait souvent à partir de l'écran du bureau vers d'autres appareils, sans tenir compte de l'environnement physique et des activités de l'utilisateur. Cette thèse explore une nouvelle forme de système projecteur-caméra à manipulation directe, qui exploite les caractéristiques uniques du mouvement physique d'une lampe de bureau pour manipuler du contenu interactif vers et depuis l'écran de bureau, mais aussi vers et depuis différents dispositifs et le reste de l'environnement physique, tout cela en conservant une connaissance du contexte de l'espace de travail.

Trois projets explorent la conception, le prototypage et les facteurs humains associés à un système de lampe augmentée dans lequel la lampe fonctionne comme un dispositif d'entrée et de sortie reliant l'ordinateur de bureau et l'environnement physique. Dans le premier projet, un espace de conception d'interaction est introduit pour la manipulation physique directe à l'aide d'une lampe d'architecte avec une preuve de concept utilisant un projecteur et un système de suivi des mouvements. Nous démontrons son potentiel à travers trois scénarios, décrivons les résultats de l'étude évaluant son potentiel et détaillons les implications en termes de conception. Dans le second projet, nous étudions les impacts sur la performance de l'utilisateur et les stratégies d'interaction lorsqu'il interagit avec une lampe augmentée dans un espace de bureau. Nous menons une expérience contrôlée en réalité virtuelle pour comprendre l'impact de deux mécanismes de contrôle pour les tâches d'acquisition de cibles avec un affichage dynamique de type peephole: "couplé", lorsque le centre de l'affichage est utilisé pour la sélection et "découplé", lorsque la sélection est gérée par des entrées séparées comme le toucher direct. Nous constatons que les deux mécanismes de contrôle présentent des différences subtiles en termes de temps de réalisation de la tâche et d'erreur, mais que les utilisateurs suivent des stratégies différentes pour coordonner le mouvement de l'écran dynamique avec les différentes techniques d'acquisition de cibles. Dans le troisième projet, nous explorons cette observation dans un contexte plus général. En utilisant un environnement de réalité virtuelle contrôlé, nous menons une expérience pour étudier si ce que les utilisateurs ont l'intention de faire avec une cible virtuelle a un impact sur la façon dont ils planifient et effectuent l'acquisition initiale de la cible. Nos résultats permettent de comprendre les profils de mouvement de l'utilisateur avant l'acquisition pour différentes interactions prévues avec la même cible. Nous discutons de la manière dont ces profils de mouvement peuvent ensuite être utilisés pour améliorer la conception de la lampe, par exemple en y intégrant des capteurs de force pour améliorer la prise en compte de l'activité de l'utilisateur. L'ensemble de ces résultats sont prometteurs pour étendre l'informatique de bureau actuelle à l'environnement physique du bureau, sur la base d'une compréhension plus approfondie des activités de l'utilisateur dans cet espace.

ACKNOWLEDGMENTS

As cliché as it may sound, none of this would have been possible without the support of so many.

An immense thank you to my supervisors, Géry Casiez, Edward Lank, Sylvain Malacria, and Daniel Vogel. My supervisors have been arguably the most impactful people in my life after my parents. Their creativity in developing novel ideas, and rigour in conducting scientific research have shaped me into the researcher I am today. Personally, I've learned to combat procrastination, refine ideas through iteration, avoid reinventing the wheel, and appreciate attention to detail and aesthetics. I'm extremely grateful for their support, guidance, trust, and patience, which have helped me grow not only as a researcher but also as an individual (bear me one more time for writing a length paragraph). To Ed, my research journey would not have been possible without you, and I will always remember the moment when we first met and talked, and your trust in me.

We all know the function of parentheses :)

To Mark Hancock, Caroline Appert, Nadine Couture, Scott Bateman, and Marcos Serrano, thank you for accepting to serve on my committee and providing insightful comments and feedback to my thesis. I would also thank Huamin Qu and Xiaojuan Ma, who brought me into the world of Human-Computer Interaction.

To all my wonderful advisors, colleagues and friends in the Loki team in Lille and the HCI lab in Waterloo, thank you for making this journey so much fun and memorable. To Keiko Katsuragawa, Anastasia Kuzminykh, Bahareh Sarrafzadeh, Rina Wehbe, Jay Henderson, Jeff Avery, Damien Masson, Hanae Rateau, Futian Zhang, Nikhita Joshi, Johann Wentzel, Matthew Lakier, Alé Luz, Ludwig Wall, Antony Irudayaraj, Xizi Wang, Nabil Hannan, Louis Kuang, Sangho Suh, Rahul Ray, Bruno Fruchard, Mathieu Nancel, Damien Pollet, Aurélien Tabard, Grégoire Richard, Axel Antoine, Alice Loizeau, Philippe Schmid, Timo Maszewski, Raphaël James, Suliac Lavenant, Eva Mackamul, Raphaël Perraud, and Felipe González, thank you for your all help, talks and discussions. Best wishes to you all.

To Yucang and Yuanhang, I am amazed and grateful that we have been friends for almost half of my life time. Thank you for being there, even we've all changed over the years. I would also like to extend my appreciation to others over the years: Ji Xin, Yangtian Zi, Zhongling Wang, Peng Shi, He Bai, Xinyu Zhang, Kaisong Huang, Shangzhe Wu, Chunyan Bai, Xinnan Du, Gengmo Qi, Haoye Cai, Boyu Liu, Wenhao Zhang, and many others. I am deeply grateful for their presence in my life. Last, I would like to thank Tim Hortons for offering coffee every single day.

To my parents, thank you for your unwavering support and love, which I could never fully repay. I hope that my life journey brings you pride and joy.

Lastly, to Meixin, thank you for healing my soul, and sharing all the ups and downs in my life.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Stay Hungry, Stay Foolish.

– Steve Jobs

1

INTRODUCTION

Desktop computing, which emerged in the late 1970s with the introduction of personal computers, has been one of the dominant computing environments for human-computer interaction for decades [53, 97]. It typically involves a stationary setup with a monitor, keyboard, and mouse, and allows users to manage various applications to accomplish diverse tasks, revolutionizing personal and professional productivity.

In 1991, Weiser envisioned a world of ubiquitous computing, where computing devices of different sizes (as in "tabs", "pads" and "boards") are *"everywhere, not just in our homes and offices, but also in our cars, our clothes and even our bodies"* [240]. Over the decades, we have seen this vision and accompanying concepts realized in the forms of smartwatches, smartphones, tablets, and laptops. Unlike desktop computing that stays in one place, these devices go with us everywhere and work together, enabling us to access information anywhere and anytime. We can switch between devices based on our needs in different contexts, allowing for more flexible interactions. The ubiquitous computing paradigm also presented new possibilities for desktop computing to enable information access from multiple devices and displays within the same space to accommodate different interaction needs.

Early work by Grudin highlighted the importance of multiple monitors in expanding a digital workspace for desktop computing [94]. Subsequent studies [26, 109] explored how users adapt their workflows to larger and multiple display setups, indicating a shift towards more flexible computing environments. As technology advanced, the focus expanded beyond static monitor and display setups, moving towards the use of diverse mobile devices, such as smartphones and tablets as external tools to access and manage information across the desktop environment [105, 187].

While the progression of desktop computing, from a single display to multi-device environments, aligns with our rapidly expanding digital lives, its cost cannot be ignored, such as the cognitive burden of switching tasks between devices and displays [6, 187] and the practical constraints of optimizing device placement on limited desk space [264]. This has led to considering how computing can be better embedded in our physical environment [23, 159], not just across devices and displays in the workspace [39].

In desktop computing, the workspace is concrete: a physical desk with a computer, monitor, keyboard and mouse. Under ubiquitous and pervasive computing, it is a more distributed and adaptable computing environment that extends across devices and physical environment.

Building upon ubiquitous computing, pervasive computing presumes an altogether different vision [19]: our interaction should be transformed from a device-centric, application-driven model to a ubiquitous, task-oriented environment where computing is embedded into our environments, emphasizing context-awareness and adaptability [202]. This offers a different direction and potential challenges to desktop computing by proposing a shift from device-centric to task-oriented and environment-aware computing. However, desktop computing by nature creates a virtual space for users to manage and perform

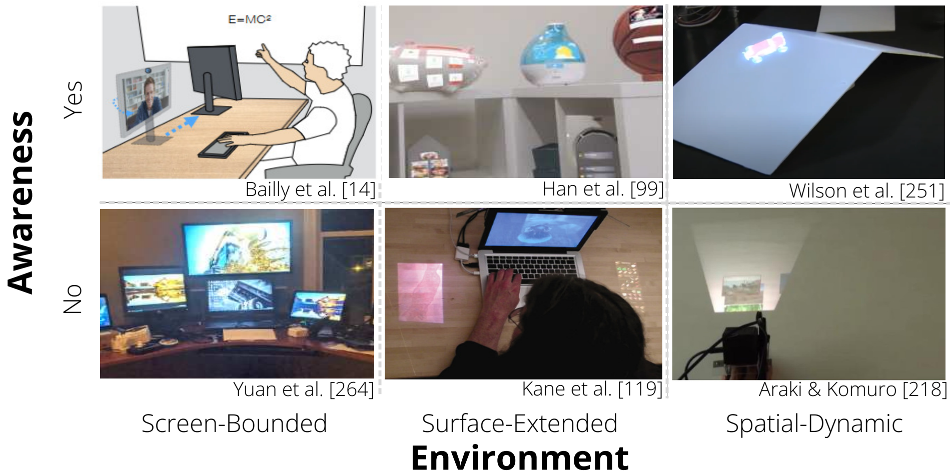


Figure 1.1: The dimensions of pervasive desktop computing based on Environment and Awareness.

tasks, rather than integrating computing into the physical environment [219, 229]. Besides, desktop computing typically has minimal understanding of users' activities, relying on additional information from other devices and systems [2, 244]. Despite these inherent limitations, there is a strong desire to evolve desktop computing toward a more pervasive paradigm, which stems from the potential benefits of merging the power and familiarity of desktop systems with the flexibility and context-awareness of pervasive computing, potentially enhancing productivity and user experience [202].

Many approaches and technologies have been explored to integrate pervasive computing into desktop environments. I propose to classify these technologies into two dimensions based on two key characteristics of pervasive computing: *Environment* and *Awareness* (see Figure 1.1).

Environment defines where users access the information within the workspace. It ranges from three dimensions: screen-bounded, surface-extended and spatial-dynamic. Screen-bounded means that the digital information stays within the physical screen even when one or more than one device and display are connected to the desktop system (*e.g.* a single monitor setup [14], multi-monitor setup [264] or multiple mobile devices [105]). Surface-extended, on the other hand, allows information to go beyond the physical screen and is extended in a static augmented physical environment (*e.g.* using a projector to display information on a table [241]). Spatial-dynamic allows surface-extended information to be dynamically repositioned to different physical surfaces and objects (*e.g.* moving an augmented mouse to manipulate a virtual menu[218]).

The other dimension, *Awareness* defines the ability of the system to understand the physical environment, and adapt content based on a user's activities. Examples include a self-organized desktop setup based on the location and hand gestures of a user around the desk [14] or context-aware interface layout based on current task loads [140].

While existing approaches have made significant strides in making desktop computing pervasive, they often address only a subset of challenges or remain confined to specific points along those *Environment* and *Awareness* dimensions. Prior work employed projectors and cameras to extend desktop computing on physical surfaces around a desk (surface-extended, non-awareness) [167, 190, 241]; or augmented traditional desktop workstation with a monitor, a keyboard or a mouse and dynamically changes devices' configuration based on user activities (screen-bounded, awareness) [14, 262]. Besides, the *Environment* spectrum of expanding beyond screens has predominantly been unidirectional [101, 115, 130, 137, 153]: they typically focus on moving content outward—from screen-bounded to either surface-extended or spatial-dynamic environments. This overlooks a crucial aspect: the natural and intuitive return of content to the original *Environment*, enabling *bidirectional* connection. In particular, moving digital content out into the physical environment from the desktop and back again remains unexplored, or or at least not obvious in prior work. For instance, a virtual toy car can adjust its pose when it hits a person or a wall in the physical environment but it cannot run across the desktop display or stay within it [251]. While information can also be embedded onto physical objects [99], it is not clear how these information can be brought back to the display as users might have different interaction needs in different contexts. In short, the bidirectional connection describes the ability of a system that moves information across these dimensions while preserving the awareness during this movement.

All together, these lead to our first research question (RQ):

RQ1 *How can we design a system or solution to enable bidirectional connection across the Environment?*

To answer this question, we design and prototype an augmented lamp, which enables bidirectional connection through its direct manipulation. Our augmented lamp is based on an architect lamp, a common desk object that aligns with our goal of embedding computing into the physical environment. Its mechanical structure allows for rich manipulation and maintains its position when not in use. This design enables understanding of users physical activities and allows users to switch contexts to meet their interaction needs (*i.e.* bidirectional connection) through the lamp's physical manipulation. Then, when manipulating such a system:

"Direct manipulation" refers to physically interacting with the lamp by touching or moving it. This contrasts with indirect manipulation, where gestures are used to control the lamp without physical contact [139].

RQ2 *How can we interact with the digital information and physical objects and space across Environment with Awareness?*

RQ3 *How does this system improve existing desktop experiences?*

The augmented lamp creates a dynamic peephole display in the physical space, and users can interact with the virtual content with different inputs. Potential impacts between the dynamic peephole display and different inputs on user performance and interaction strategies are unclear, more specifically:

RQ4 *How do "coupled" control, when the display centre is used for selection, and "decoupled" control, when the selection is handled by separate inputs like direct touch, impact target acquisition tasks in a dynamic peephole display?*

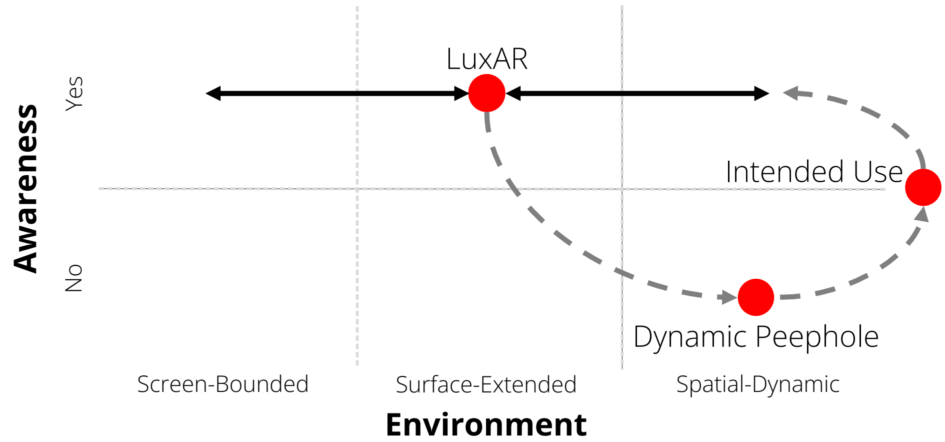


Figure 1.2: Solid lines with arrows indicate how our work expands across dimensions, and dashed lines with arrows indicate how our work connects within dimensions.

RQ5 *How do different interaction techniques impact user's physical manipulations?*

1.1 RESEARCH OBJECTIVES AND OVERVIEW

We examine these five questions from three projects. The proposed system and related user studies are represented in Figure 1.2. The research questions for each specific project, the overview, and flow of the research questions are presented in Figure 1.3. Simply put, the high-level research objective of this thesis can be stated as:

This thesis is a product of collaborative research, and I use "we" when discussing the projects to highlight the contributions of my supervisors.

> *Design an augmented lamp that makes desktop computing pervasive, and understand the impacts induced from its design on user behaviors*

Below, we summarize the steps we took to address these research questions.

1. To answer RQ1, we propose LuxAR, an augmented architect lamp on a desk as an input and output system to extend and augment the desktop computing onto physical environment. It features a pico-laser projector for output, a button for direct input, and camera tracking for real-time position and orientation. This enables moving and display content bidirectionally across physical displays, devices, and surfaces, akin to mouse input in desktop computing.

2. To answer RQ2, we explore the design space of an architect lamp through its unique physical characteristics: flexible manipulation and stable positioning after manipulation. We purpose different interaction techniques to enable bidirectional virtual content flow across the *Environment*, accommodate inputs commonly found on the desk, understand and augment the physical environment.

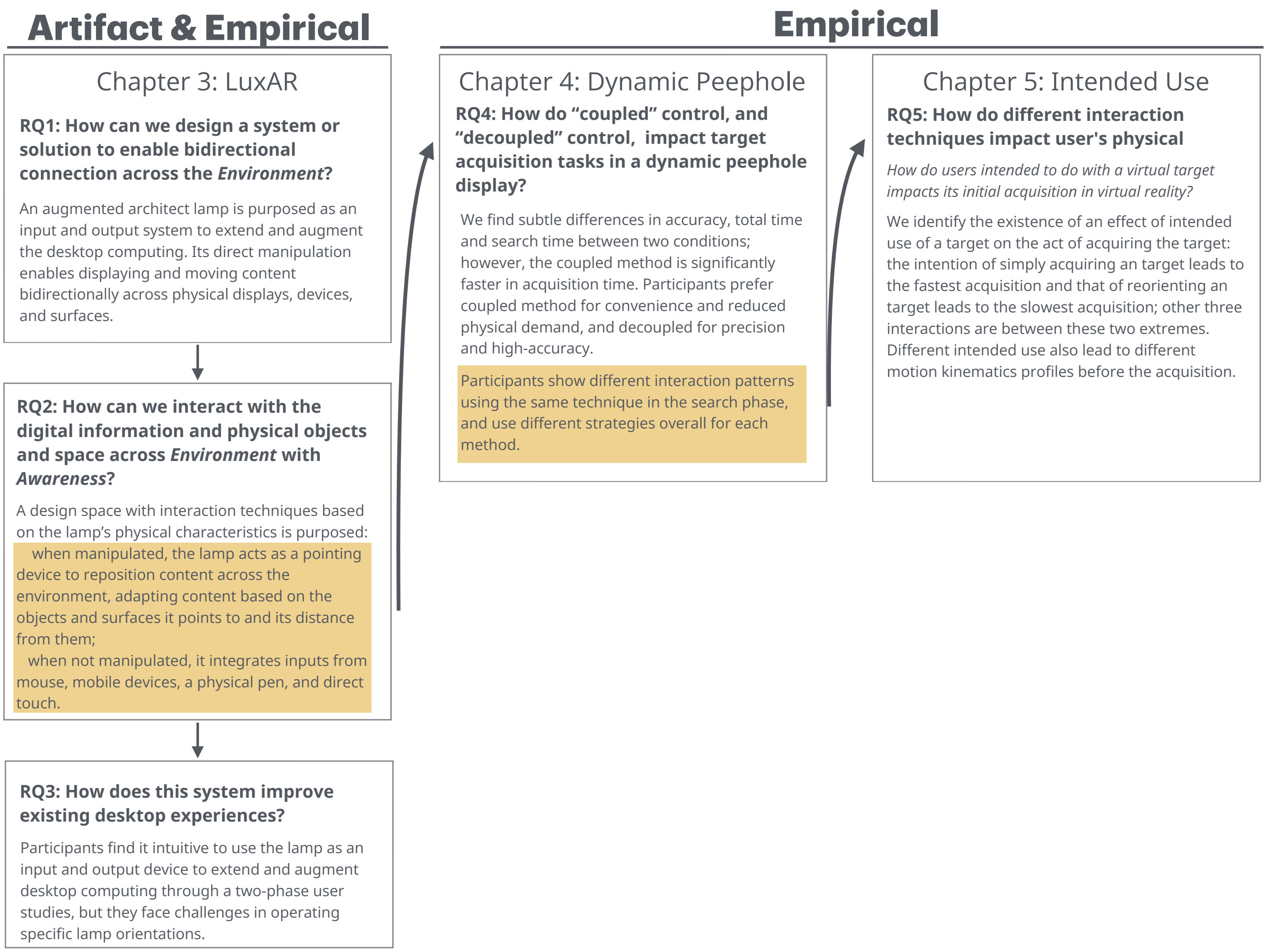


Figure 1.3: Chapters with main research questions (bold), specific research questions (italic), main outcomes, research path connecting each chapter and horizontal text highlighting the type of contributions.

3. To answer RQ3, we purpose three scenarios: a calendar, a music player and a drawing annotation tool, and conducted a two-phase user studies to evaluate the proposed interaction techniques and understand how the augmented lamp can improve the existing desktop experiences.

4. To answer RQ4, we simulate the dynamic peephole problem on a virtual desk and conduct a controlled experiment of 2D target acquisition task with coupled and decoupled control mechanisms. We aim to understand quantitative differences and collect qualitative feedback of these methods. Results reveal further design consideration for inputs and interaction techniques for RQ2 and applicable scenarios in RQ3.

5. Results from RQ4 suggest different interaction strategies impact initial target acquisition. To answer RQ5, we explore this observation in a more general context. We conduct an experiment in a controlled VR environment

to investigate whether what users intend to do with a virtual target impacts how they plan and perform the initial target acquisition. We analyze their quantitative differences in terms of accuracy, time and motion kinematics measures.

To summarize, in this thesis, we explore the design space on the physical manipulations of an architect lamp, build the prototype of the system, purpose interaction techniques, and conduct user studies and control experiments to understand the human factors involved in a pervasive desktop computing environment.

1.2 CONTRIBUTIONS

We summarize the research contributions by project, with high-level HCI contributions type [254]. For each, we outline the key results and insights that form our contributions.

1.2.1 *LuxAR: A Direct Manipulation Projected Display to Extend and Augment Desktop Computing*

In Chapter 3, we prototype and evaluate a desktop input and output device in the form of an architect desk lamp to create a pervasive desktop computing environment. We present a set of interaction techniques based on the direct manipulation of the lamp where virtual content can be transferred across displays, devices and the surrounding physical environment and the representation of content can be adapted through its direct manipulation (*Artifact Contributions*). A two-phase semi-structured user studies with the prototype evaluate the proposed interactions and consider potential scenarios and applications. Based on the results, we propose further design considerations for direct manipulation systems to extend and augment desktop computing (*Empirical Contributions*).

1.2.2 *Investigating Coupled and Decoupled Target Acquisition of Dynamic Peephole on an Augmented Desk*

In Chapter 4, we conduct a controlled experiment in a virtual environment investigating how coupled and decoupled control mechanisms impact target acquisition in dynamic peephole interaction on an augmented desk. We aim to understand quantitative differences and collect qualitative feedback of these methods. Results show subtle differences in accuracy, total time, and search time between two methods for the task, but the coupled condition is faster in acquisition time significantly. Participants preferred each condition for different reasons. We also find that using the same technique for searching targets, participants tend to move the peephole faster with the coupled condition but reveal targets later, and move the peephole slower but reveal a target earlier with the decoupled condition. Participants also demonstrate different interaction strategies with each condition. Our findings suggest that different methods may be advantageous for certain tasks and scenarios in augmented desktop environments, informing future design guidelines for dynamic peephole interfaces (*Empirical Contributions*).

1.2.3 *Exploring the Effects of Intended Use on Targeting in Virtual Reality*

In Chapter 5, we conduct a controlled experiment with five intended manipulation tasks for a target in a virtual environment: targeting, dual-targeting, throwing, docking and reorienting, and investigate their impacts on the target's initial acquisition from a quantitative perspective in terms of accuracy, time, and motion kinematics measures. Our results demonstrate that the intended use of a target affects its acquisition time, and correspondingly, the movement towards the target, including the peak velocity and time to the peak velocity. We contribute to an understanding of the intended use impact, which can be applicable to a wide range of tasks (*Empirical Contributions*).

1.3 THESIS OUTLINE

The remainder of this thesis is organized as follows.

In Chapter 2, we review the literature on extending and augment desktop computing to diverse devices, physical objects and physical environments, with a focus on Spatial Augmented Reality (SAR) methods. We discuss the control of projected display of SAR methods and interaction methods with both the virtual content and physical environments.

In Chapter 3, we describe a system prototype to extend and augmented desktop computing, its design space on the physical manipulation of an architect lamp, applicable scenarios, and user studies to understand and evaluate the prototype.

In Chapter 4, we conduct a controlled experiment to understand the differences between two control mechanisms on dynamic peephole target acquisition and their impacts on interaction strategies.

In Chapter 5, we conduct a controlled experiment to study the impacts of intended use of a target on its initial acquisition.

In Chapter 6, we summarize contributions, highlight potential impacts, summarize limitations, discuss future opportunities, and mark final word.

2

RELATED WORK

In Section 2.1, we present how prior methods and systems make desktop computing pervasive and explain why our method uses Spatial Augmented Reality (SAR). We categorize control methods of projected display in SAR into three categories in Section 2.2. Then we discuss interaction methods that enable users to engage with digital information with different control methods in Section 2.3. Finally, in Section 2.4, we present interactive experiences in the physical environment that come with the use of SAR. Specific related work is reviewed in subsequent chapters.

2.1 MAKING DESKTOP COMPUTING PERVASIVE

To understand how pervasive computing paradigm can be applied to traditional desktop environments, we turn to Abowd and Beale's framework for interaction in computing environments [1]. This framework describes a cyclical process in which users provide inputs to perform tasks in a system, which then presents outputs observed by users. We relate this framework to our dimensions (Figure 1.1), where *Environment* represents the space in which users perform inputs and receive outputs, and a system has *Awareness* to sense inputs in the *Environment* and adapt outputs accordingly.

2.1.1 *Screen-bounded but Pervasive Desktop Computing*

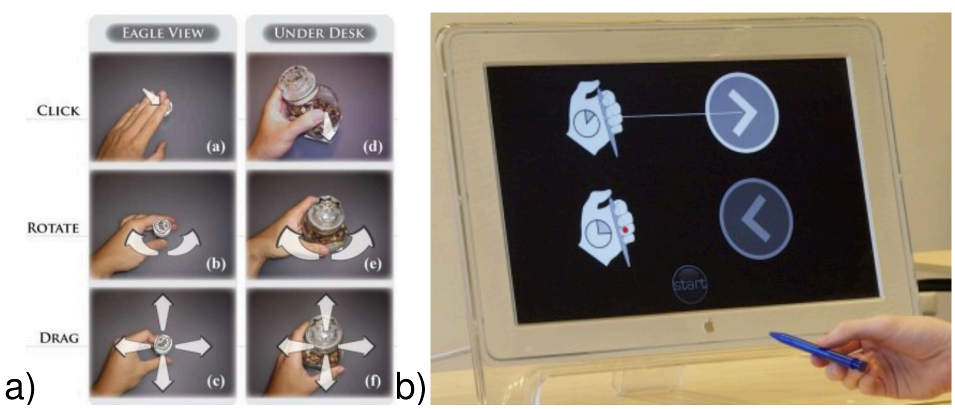


Figure 2.1: Examples of interacting with digital information using everyday objects by a) manipulating their poses [60] (*e.g.* rotating a bottle); or b) semantics mapping between interfaces and physical actions on them [65] (*e.g.* pushing a pen cap for a button control).

Even with the confines of physical screens, desktop computing can be made pervasive by augmenting the physical environment's input capabilities. One prominent method to achieve this goal is the use of tangible user interfaces (TUI)[110, 111, 229] which aims to manipulate digital information through physical objects. The metaDesk [230] is an early demonstration of this concept, allowing users to interact with geographical information using "phicons" (physical icons). Building on this concept, researchers have explored various applications of TUIs for desktop computing. Fails et al. find that using physical objects as input is valuable due to greater engagement, enhanced interaction and better learning outcomes for children [79]. For high-stakes environments like control rooms, Müller et al. demonstrated how physical controls can enhance user interaction, reduce cognitive load, and improve overall system efficiency [163]. Customized physical objects have been used as input to enhance spatial awareness in people through 3D content manipulation tasks [88, 131] and assist people with visual impairments to interact with digital content [16] on the desktop. Besides customized objects, everyday objects have also been explored for this purpose. Cheng et al. place fiducial markers on physical objects such as bottles, use cameras to track the manipulation of these objects, and enable people to toggle controls in desktop applications, such as controlling a button by rotating a bottle cap [60] (Figure 2.1 (a)). The instant user interface, proposed by Corsten et al., explores the concept of allowing people to assign different meanings to various poses or touches on physical objects and transforming everyday objects into function input devices for interactions [65] (Figure 2.1 (b)). An example is the use of furniture, such as chairs [180, 181], which can be used to control desktop applications through people's poses while being seated.

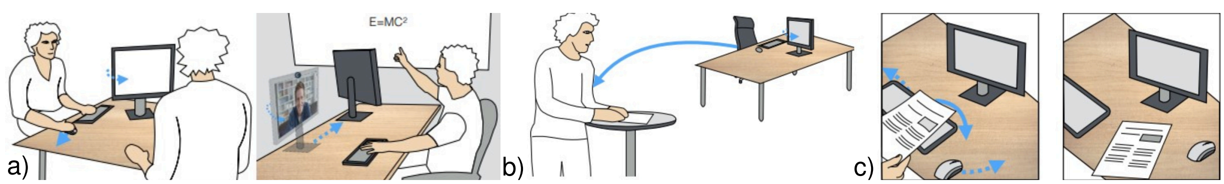


Figure 2.2: The LivingDesktop system [14]: mouse, keyboard and monitor in the workspace can be automatically configured based on the users' activities: a) sharing a monitor with co-worker using keyboard or pointing to an object; b) the monitor follows the user for consistent reading experiences; c) the system automatically adjust components' positions to make room for an object.

Another approach to making screen-bounded desktop computing pervasive is by detecting users' activities through body movements using sensing and tracking technologies. The LivingDesktop system [14] developed by Bailly et al. integrates sensors, motors and vision-based systems to detect user activities around the desk, allowing components like the mouse, keyboard and monitor to move autonomously to improve ergonomics and facilitate collaboration (Figure 2.2). Similarly, the DeskWave system [152] integrates microwave arrays beneath the desk to sense user activities above the desk. Researchers have also used desktop cameras to predict intended keyboard shortcuts of users based on finger postures [267] or planned actions based on hand postures [62], further enhancing the interaction between users and their digital workspace. Meanwhile, mobile devices like smartphones, smartwatches, and tablets, equipped with diverse sensors, have been explored to extend desktop computing capabilities [39, 105, 184]. However, these devices often function as additional displays for desktops and standalone computing units, with information largely confined to their small screens. While they possess rich sensing capabilities, these are primarily used for themselves rather than to understand the physical environment or users' activities in ways that could augment and expand desktop computing.

2.1.2 *Going Beyond Screen to Static Physical Environment*

The Use of a Fixed Projector

To extend digital content beyond the physical screen of the desktop and into the surrounding environment, one of the most common methods is to use projectors. Projectors are typically placed at a fixed location to create a static projected display on a physical surface. This technology has its roots in the late 1980s and early 1990s when projectors began to be used for presentations [64] and remote collaboration [223]. The advancements in projectors lead to more vivid, sharp, and clear images in the projected display, creating more interaction opportunities. The Touch-display keyboard system [33] is one of such examples. It uses an overhead projector to transform keyboards into interactive surfaces, overlaying contextual information over each key, and creating a dynamic display of virtual keyboards (Figure 2.3). This greatly expands the functionality of a traditional input device. More recently, Alkayyali et al. push the boundaries even further by exploring the creation of interactive displays on walls using photochromic paint activated by laser projectors, allowing people to dynamically create customized inputs and outputs around the desktop space [5].

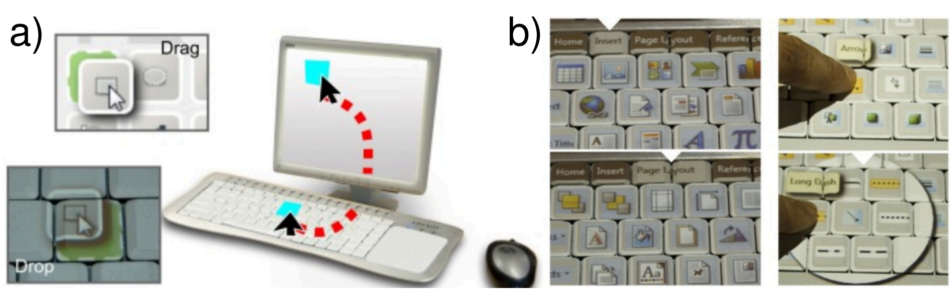


Figure 2.3: The Touch-Display Keyboards [33] overlays contextual information over each key and allows people to customize the function of each by drag&drop by mouse.

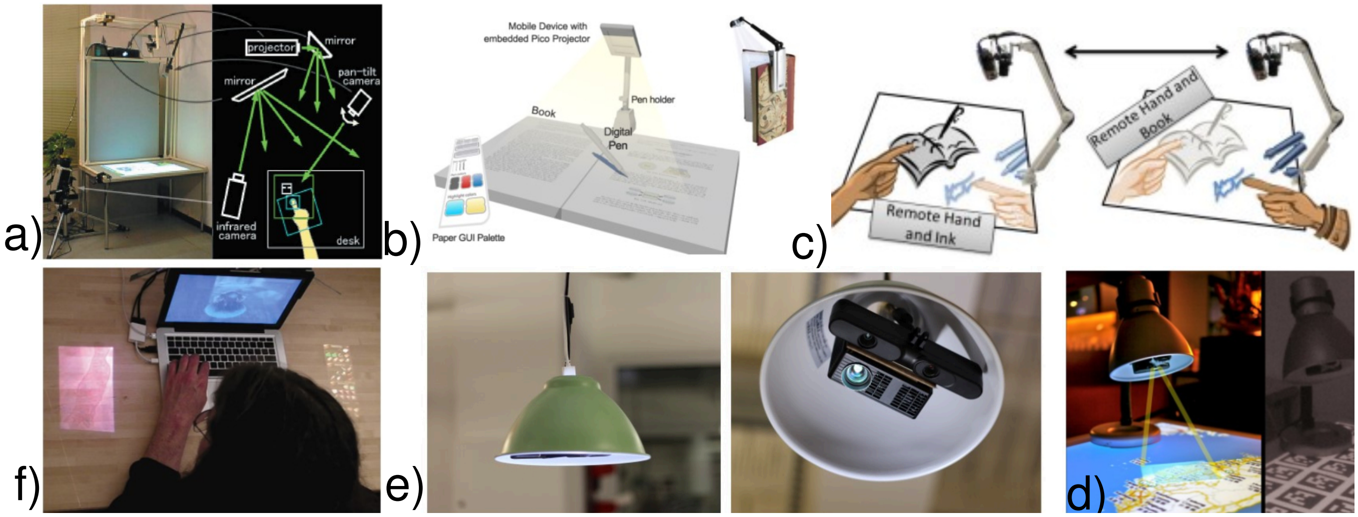


Figure 2.4: Examples of fixed projectors and cameras in clockwise direction whose setups are: a) on top the desk [130]; b) attached to a bookmark [67]; c) and d) inside a tabletop lamp [54, 117]; e) inside a ceiling-hung lampshade [259] and f) attached to a laptop [119].

The Use of Fixed Projectors and Cameras in SAR

The integration of cameras with projectors marks another significant step forward, enabling not only the output of digital information onto physical surfaces, but also input directly from the physical space, enabled by hand and body movement tracking. One of the earliest examples is the DigitalDesk system [167, 241, 242]. This system moves multiple applications, such as a calculator, from the desktop to physical paper and enables users to use their fingers for interactions. The idea of not only moving digital content on physical environments and objects, but also enabling interaction with them, is commonly referred to as Augmented Reality (AR) [147, 243], and the use of projectors and cameras forms the basis of Spatial Augmented Reality (SAR) [31], which encompasses various techniques for merging digital and physical spaces. The concept of SAR originated from *the future of office* [190]. It demonstrated the potential of SAR by placing multiple projectors and cameras in the ceiling to overlay digital information directly onto physical objects and surfaces through projection mapping without the use of markers. A huge amount of works adopted similar setups as *the future of office* to offer room-sized interactive experiences [24, 114, 115, 173, 250]. Some works placed this pro-cam unit on everyday objects, such as tabletop lamps [54, 124, 138] (Figure 2.4 (d)), a bookmark [67] (Figure 2.4 (b)), or laptop [119] (Figure 2.4 (f)) to provide digital ink on physical paper or facilitate remote collaboration [117] (Figure 2.4 (c)) in the desktop computing environment. Other researchers explored marker-based approaches, such as the Augmented Surfaces system [192], which placed a projector and cameras on the ceiling, and used fiducial markers on physical surfaces and devices to allow the sharing of virtual information through the projected display on the desk. The EnhancedDesk [130] employed a similar approach to overlay digital information on physical books (Figure 2.4 (a)). Given the nature of SAR systems, they have been designed to be environment-aware, adapting projected content based on the physical environment and user actions. Riemann et al. proposed a responsive layout optimization method for the virtual content based on the geometric information of physical objects [194]. Xiao, Hudson, and Harrison demonstrated how virtual elements can automatically rearrange themselves to avoid overlapping with physical objects on a messy desk, creating a more seamless integration between digital and physical spaces [259] (Figure 2.4 (e)).

The Use of Head-Mounted Display in AR and MR

An alternative approach to extending digital content beyond the desktop screen is the use of head-mounted displays (HMDs) for Augmented Reality (AR) and Mixed Reality (MR), such as Microsoft Hololens and Meta Quest 3. Virtual content can be placed directly around the desktop screen [36, 128, 153, 266] (Figure 2.5 (a)). Unlike fixed SAR systems, HMDs offer great flexibility in terms of static positioning. In addition to being displayed around the physical screen, content can be positioned in the physical space around people [141] or anchored around physical documents or surfaces [137] (Figure 2.5 (b)). Han et al. recently propose to blend virtual information onto physical objects based on their geometry characteristics [99] (Figure 2.5 (c)).

Figure 2.5: Examples of digital information through HMDs are statically a) anchored around a desktop screen [128]; b) placed around physical documents [137]; and c) blended into physical objects [99] [a].

[a]

2.1.3 *Going Beyond Screen to Dynamic Physical Environment*

Some researchers have explored more dynamic approaches that allow digital content to move and adapt within physical environments, rather than fixed SAR setups. These systems often employ steering control mechanisms or leverage human body movement to enable dynamic repositioning of digital content.

The Use of Steerable SAR

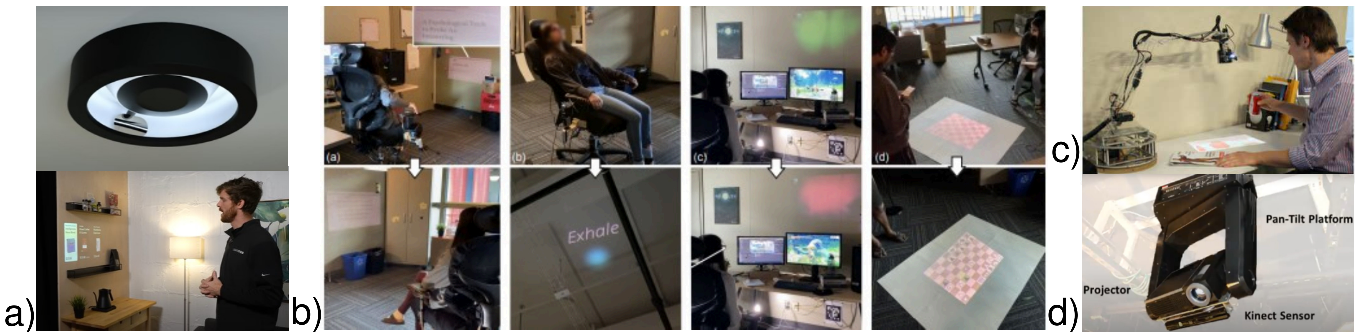


Figure 2.6: Examples of steerable SAR move virtual information by a) adjusting the position of a mirror [182]; b) changing people's postures on a chair [116]; c) tracking people's hands [139]; and d) following people's gestures and speech [251].

The steerable SAR either adjusts the projection while keeping the unit stationary (*e.g.* a steering mirror) or adjusts the pose of a projector directly (*e.g.* a pan-tilt platform). These systems commonly use cameras to detect activities in the physical environment and reposition the content accordingly. For instance, while both the Everywhere Display [177] and Escritoire [13] systems use a pair of fixed projector and camera setup and a mirror to direct projected content onto physical surfaces, the former implements a computer-controlled mirror to dynamically reposition the projected display in the surrounding environment of the desktop space. One recent example is Project LFX [182]. It is a ceiling-light platform that tracks people's locations in a room and rotates a mirror to reposition the content, which is then adapted to the physical context (Figure 2.6 (a)). Prior systems can also directly adjust the orientation of a projector. The Beamatron system [251] mount a steerable projector on the ceiling, enabling people to use speech and gesture to direct the projected content onto different physical surfaces (Figure 2.6 (d)). Joshi et al. install this type of setup on a chair so that the content can be re-positioned on the ceiling, walls or even drawers, and adapted based on people's postures on the chair when they are sitting in front of the desk [116] (Figure 2.6 (b)). The LuminAR system [139] integrates a projector and a camera into a robotic lamp, and allows people to control where and how information should be displayed by hand gestures on the desk, creating a highly flexible and adaptive display system (Figure 2.6 (c)).

The Use of Handheld SAR

In addition to a steering implementation, existing systems have explored using handheld implementations to dynamically augment physical surfaces. This method often uses projection mapping methods or reflective markers that are tracked by motion cameras. For instance, the iLamps system [188] is proposed to address dynamic projection issue on non-planar surfaces, such as room corners to display virtual content. Besides simply using this design to display content, the use of a flashlight or spotlight metaphor enhances these systems beyond output devices and transforms them into input devices [186, 200]. Based on this idea, Beardsley et al. propose using a handheld projector as a pointing device for selecting content [22] (Figure 2.7 (a)). Blasko, Coriand, and Feiner propose a wrist-worn SAR system in which people can move web pages on the wall with their forearms and pan the content with their wrists [32] (Figure 2.7 (b)). The tilt-based manipulation can also be used to zoom in on projected content and take snapshots when reading physical documents [118]. Leveraging the six degrees of freedom of a handheld SAR system, Cao and Balakrishnan integrate this handheld device with a tracked pen, enabling users to customize interactive displays and annotate both virtual and physical content in a room [45] (Figure 2.7 (d)). Willis, Poupyrev, and Shiratori propose MotionBeam, a framework that describes the projected content's interactions through the physical manipulation of a handheld projector system [249] (Figure 2.7 (c)). For instance, people can tilt the system to adjust the viewing perspectives of displaying the projected content. They later introduced HideOut [247], an innovative system using IR-absorbing ink to create hidden fiducial markers on paper documents such as storybooks or board games. By moving their handheld systems and dynamically changing the displayed content, people could bring these documents to life, enabling immersive storytelling and gaming experiences. This use of handheld SAR systems also created an opportunity to engage multiple users for collaborative and shared experiences [46, 66, 248]. One variant of handheld SAR is to integrate a pro-cam unit into physical objects, such as a mouse [9, 218] (Figure 2.7 (e)) or a lamp [135] (Figure 2.7 (f)) that require physical manipulation. These diverse approaches demonstrate the potential of handheld and wearable SAR systems, each offering unique ways to interact with digital content in physical spaces. Meanwhile, these devices must be held in the air for aiming and input movements, fully occupying at least one hand and causing fatigue.

The Use of HMDs

HMDs also enable dynamic content placement in physical spaces, offering yet another approach to pervasive computing. Most common approaches employ controllers or hand gestures, such as pinch, to pick and drop information at different locations [8, 161]. Mcgill et al. conduct studies investigating methods for the arrangement of virtual windows and find that the use of head orientation to control the position of the display have significant benefits in the seated workspace [153]. Other researchers have explored more sophisticated approaches to content placement. For instance, Lindlbauer, Feit, and Hilliges present an optimization-based method to automatically control when and where information appear, and how many details it presents to people based on people's current activities. Cheng et al. propose optimizing the placement of digital information based on semantic relationships with physical objects,

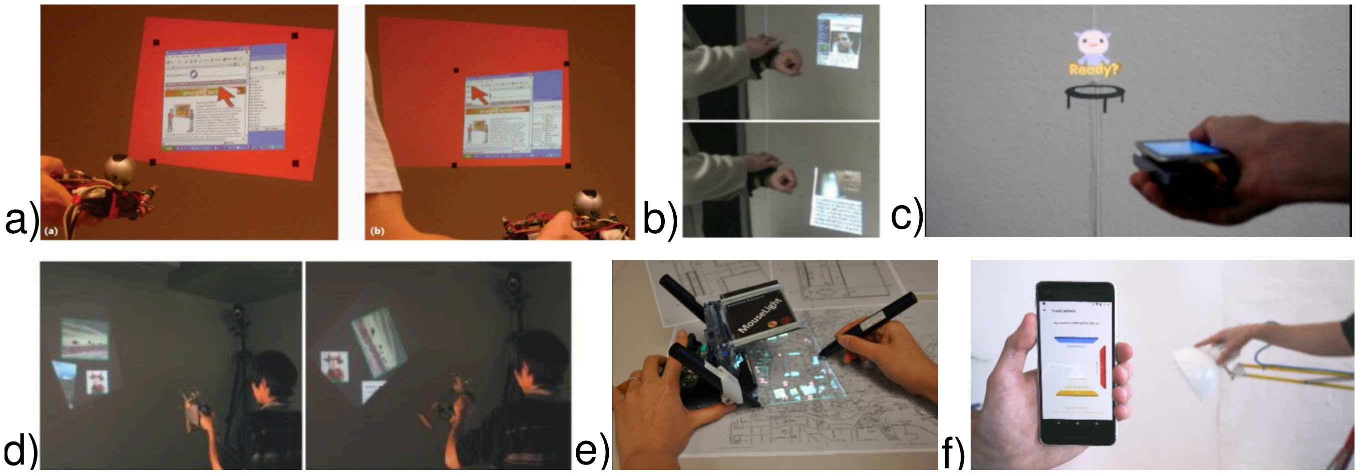


Figure 2.7: Examples of handheld SAR to dynamically move virtual information onto different physical surfaces, so people can a) point to it using a projector [22]; b) zoom into it by rotating their wrist [32]; c) place it on a physical object to create animation [249]; d) moving and rotating it by adjusting the pose of a projector [45]; e) overlay and move information on a physical document using mouse [218]; and f) display different information on surfaces by adjust the lamp's head [135].

creating a more context-aware augmented environment [61]. Recent work has even explored allowing users to dynamically optimize content placement through gestures [168], giving users more direct control over their augmented environments (see Figure 2.8).

2.1.4 *Moving to Bidirectional Pervasive Desktop Computing with SAR*

Pervasive desktop computing has shown significant improvements in various dimensions. The benefits of pervasive desktop computing are evident in its enhanced interactivity and improved efficiency, building on top of the user familiarity of the traditional desktop computing.

However, despite these improvements, current methods either focus mainly on addressing a specific problem, or neglect the bidirectional connections between *Environment* dimensions. Screen-bounded systems either use physical objects [60, 65] as inputs or track users' activities in the space for interacting with the digital information[62, 152, 267]. However, these approaches confine digital information within screens, failing to extend it into the physical space. Other methods, employing projectors or HMDs, move content beyond physical displays, and enable static augmentation [33, 99, 119, 137] or dynamic augmentation [45, 116, 135, 168, 182] on physical surfaces and objects. However, they often lack the flexibility to adapt to varying user activities, and interaction contexts. For instance, consider a sticky note application: it could be moved from the desktop to the physical desk for easy examination, then to a wall for reminders, or onto relevant objects for contextual association. Importantly, it should also be movable back to the desktop for editing when needed. Existing systems in the *surface-extended* dimension often fall short in

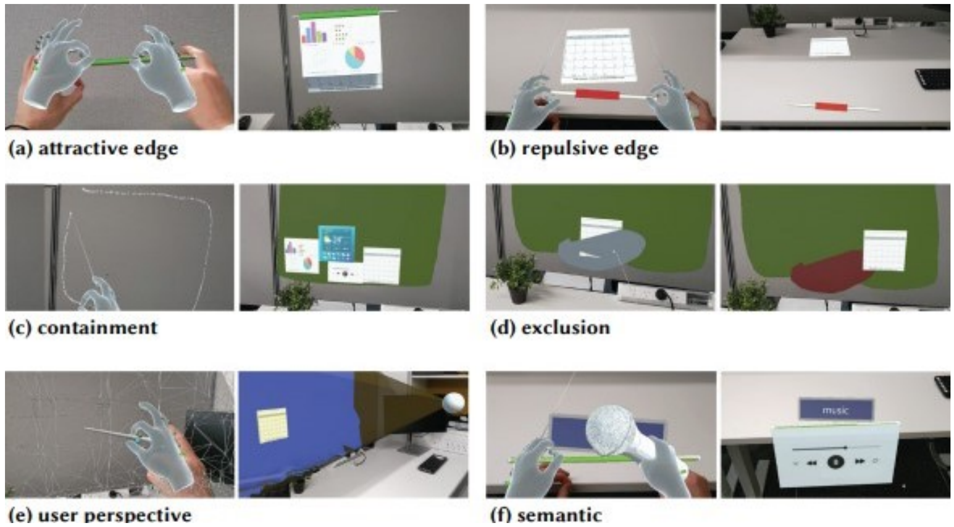


Figure 2.8: Wearing an MR HMD, people can perform different gestures to define where and how content should be placed in the physical environment [168].

different aspects. For instance, some systems with a pair of fixed pro-cam setup [54, 119, 124], focus on a specific surface without the ability to transition to other surfaces or objects. Some works can adaptively augment objects [99, 194, 259] but lack mobility. Others cover multiple surfaces for augmentation [24, 115, 173] but neglect physical object augmentation. Meanwhile, while methods in *spatial-dynamic* dimension enabled the virtual content dynamically positioned across different surfaces and addressed the virtual content adaptation simultaneously [32, 45, 116, 168, 182, 249], steering control systems might not consistently reflect users' intentions and needs accurately, and handheld systems or HMDs typically lack support for extended use for the desktop computing environment [160]. Most importantly, the concept of returning information to its original *Environment* dimension, such as from spatial-dynamic to surface-extended, or back to the desktop or other screens, is largely overlooked in these methods.

This thesis aims to address this gap by building a bidirectional connection across *Environment*, enabling users to interact with digital content and physical environments simultaneously. Our approach focuses on preserving awareness of physical objects and surfaces, ensuring that the digital augmentation enhances rather than obscures the physical world.

Among the various methods, SAR stands out as a particularly promising method to achieve our goals because it can support flexible and dynamic interactive displays over the physical environment, to be integrated into everyday objects, leveraging their physical properties for interactions, and support long-duration tasks that are required for a desktop computing [160], compared to the use of HMDs. In Chapter 3 of this thesis, we describe an augmented lamp as a SAR system. It transforms the physical surfaces surrounding a desk into interactive displays, presents visual content on different surfaces, devices and displays, and augments the physical objects in the environment. Its unique physical characteristics enable us to design interaction techniques to establish the bidirectional connection across *Environment* with *Awareness*.

2.2 A CATEGORIZATION BY USER CONTROL OF PROJECTED DISPLAY IN SAR

Considering these diverse configurations of SAR setups and systems, we further categorize them based on the kind of user control over an interactive display on physical surfaces. This means how users manipulate the position and orientation of a SAR system to adjust the position of a projected display to augment surfaces and present virtual content while these systems are running: *No user control*, *Implicit user control*, and *Explicit user control*.

No user control is commonly seen in conventional SAR systems (*e.g.* [101, 115, 190, 241]). They typically use fixed pro-cams mounted on ceilings or tripods, but these require multiple pro-cam setups to cover different surfaces with reasonable resolution [101, 115, 190]. Prior research has investigated ways to minimize the number of pro-cams while still supporting interactive experiences across multiple surfaces. Maeda et al. use a fixed fisheye pro-cam for omnidirectional displays, but its configuration limits projected information to areas with fiducial markers [148].

To relax such constraints, researchers explored *implicit* and *explicit* methods for controlling the projected display. Implicit control leverages contextual cues to interpret interactions [205]. For instance, Project LFX [182] and Beamatron [251] track a user in a room and implicitly reposition projected content onto the nearest projectable surface. Joshi et al. [116] mount a pro-cam on a chair, using the chair pose to implicitly control the projected display content and location in the environment. However, implicit control relies on behavior and context analysis which may not always accurately reflect the user's intentions and needs. In contrast, explicit control means that the user decides exactly where a projected display is positioned and how it is used.

Explicit user control over SAR systems can be *indirect* and *direct*, depending on whether the projection display is manipulated physically by users. For instance, LuminAR [139]uses hand gestures to move an actuated pro-cam lamp to position a projection display on a single desktop surface; Beamatron [251] combines body poses and voice commands from users to guide the display to a target position in a room. These indirect control systems enable users to control where the projected display is located, but users have limited interactions with the content when adjusting the systems' pose. Direct manipulation increases input expressiveness and enables users to fully control all aspects of a projection display. A straightforward approach is to hold a pro-cam in the air and aim it at surfaces. Cao and Balakrishnan [45] developed a handheld SAR system to annotate virtual and physical space. By aiming a handheld pro-cam in space, it can reveal stories hidden in a storybook [247] or enable collaborative interactions between multiple users on a wall [248]. MotionBeam [249] engages users with projected anime characters through its manipulation and awareness of the physical environments. As a variation of the handheld approach, Blasko, Coriand, and Feiner [32] developed a wrist-worn SAR system for manipulating projected content, like selecting, panning, and scrolling information on a wall. A handheld system can also project information onto physical objects to interact with them, such as controlling lamps and televisions [206].

However, a notable limitation with most of these SAR systems is that interactions are *attached* to the direct manipulation. Users need to continuously carry and operate them, making it impossible to interact without holding the system in hand and easily leading to fatigue. In contrast, user interactions in implicit or indirect systems are usually *detached* to their pose and manipulation, *e.g.* users can control a virtual car across various surfaces using a joystick controller [251], or type on a virtual keyboard with both hands on the table [139], without the need to hold and manipulate the system. Besides, most direct manipulation SAR systems emphasize on interacting with content, which may lead to a lack of space awareness and adaptive interactions based on the environments the system references, including surfaces, objects, and devices.

Our system in Chapter 3 extends desktop computing through explicit and direct manipulation. It enables the user to interact with the projected augmentation by adjusting the pose of an architect lamp (*attached to the manipulation*) and when the lamp is stationary and maintains the augmentation, it allows users to interact with the augmentation through other inputs (*detached from the manipulation*). This bridges the desktop and physical spaces, incorporates multiple inputs, and facilitates adaptive adjustments of virtual content to environmental and contextual changes, establishing bidirectional connections across *Environment* dimensions.

2.3 INTERACT WITH DIGITAL INFORMATION VIA SAR

Based on our previous classification of user control over the projected display, we examine interaction methods for virtual content within the augmented surfaces, when it is more than just an image. These methods enable users to select, move and alter the content within the augmented surfaces.

2.3.1 *No User Control SAR*

For no user control SAR systems, interaction methods focus on how users engage with projected content within predefined augmented areas.

Direct touch interaction is one of the most intuitive methods for users to engage with virtual information, especially when it is within reach. Many systems incorporate infrared or depth cameras to track users' hands and fingers, enabling direct touch to select the projected content [130, 257–259]. Xiao, Hudson, and Harrison explored how touch can facilitate the cohabitation of the virtual information and physical objects in a desktop workspace [259]. They proposed ten touch gestures that allowed users to annotate virtual content based on the presence of tangible items, such as snapping a virtual menu to a book or scaling a virtual map within a book cluster.

External devices for user input are often employed when touch interaction is not feasible or desirable. These can range from traditional input devices like mice and keyboards to more specialized tools designed for specific SAR applications. In the Augmented Surfaces system [192], people can move content and exchange information between different surfaces and devices with mouse operations. This approach is particularly useful when precise input is required. Other examples of device interaction include pens, smartphones and tablets. A pen can be used to annotate virtual content directly on the surfaces [124, 126, 218] or change the content by changing its pose above the air [217]. Smartphones and tablets have been used not only as input devices but also as secondary displays, providing additional information or control interfaces [35, 121, 151, 208].

Vision-based gestures are implemented to bridge the gap between touch and external device interaction. This method is commonly seen in large-scale augmentation (*e.g.* room-level experiences [24, 114, 115, 173, 250]) where direct touch is impractical and allows users to interact with the projected content from a distance, using natural hand or body movements. A comparison study on 2D object manipulation on a projected display revealed a strong preference for a manipulation interface based on pointing gestures with tiny hand movements and minimum body movement [236]. In addition to hand gestures, shadows and palm-silhouette were also effective interaction methods. Users may pick and move the information by rotating the palm-silhouette or control the volume of sound by rotating the palm-silhouette [261].

2.3.2 *Implicit User Control SAR*

Implicit user control in SAR systems adapts their augmentation based on user behavior or environmental factors, creating a more responsive and personalized experience. While these systems also leverage body movements and user gestures to alter the projected content like what systems with no user control do, they do not explicitly ask for inputs but infer users' attentions by collecting contextual cues in the physical environments. One of the examples is the Flexible Display system [55], which changes the augmented views of artifacts based on where people stand in front of a wall. This creates an interactive museum-like experience where the content adapts to the viewer's perspective. Project LFX [182] employs a similar concept. It tracks users' locations in a room and then adjusts the projected content accordingly. When people are watching television, the system detects this activity by the user's pose and then moves the projected display over the television. It can then create an extended visual overlay to sync with the content shown on television, an experience similar to [114]. Beyond just position, some SAR systems take into account more human postures and contextual information. The augmented chair by Joshi et al. exemplifies this approach by adapting projected content based on people's postures while seated [116]. This method of implicit control allows users to receive personalized information or experiences without having to explicitly request them, creating a more immersive and calm augmented environment.

2.3.3 *Explicit User Control SAR*

Explicit user control in SAR systems require users to explicitly determine the positioning and utilization of projected displays and content. However, the direct explicit control presents different interaction possibilities from indirect one as users have to physically manipulate the system itself as input.

Indirect Control Methods

These methods allow users to manipulate the projected display and content without physically touching the system itself.

Hand gesture is the most common method to interact with the projected content. In the hat-attached system by Mistry, Maes, and Chang demonstrated this by allowing people to perform various actions with their hands to select, zoom and pan virtual content projected on distance surfaces [158]. Cowan and Li explored similar interactions in collaborative situations by the use of shadow motions [66]. The LuminAR system [139] used hand gestures to control an actuated pro-cam lamp, demonstrating interactive manipulations on both the projected display and content.

Voice command can also be integrated with hand gestures to interact with the digital content, as in the Beamatron system [251]. The system can localize people in the room but not in view of the camera unit and enable them to call the pro-cam unit to view the user. It can also allow users to point to a specific surface, like a wall and then call the unit to reposition the projected display and adjust the content.

Smart devices, such as smartphones are also used to change the pose of a SAR system. For instance, Cauchard et al. integrated a smartphone with a steerable projector so that users use the smartphone to adjust the prototype's orientation [51]. Similarly, Scheible et al. [204] and Brock et al. [38] both used smartphones as a control device to manipulate drone-attached SAR systems and interact with the projected content in the wild.

Direct Control Methods

A huge amount of work in direct control methods involves physical manipulation of the pro-cam unit or direct interaction with the projected content. Rukzio, Holleis, and Gellersen presented a survey on the use of pro-cam for pervasive computing [200]. It summarized the space of input control of using handheld systems to interact with the projected content, which could be mainly divided into two categories — direct physical manipulations and button interactions.

Hardware buttons on the systems provide an efficient way to interact with the projected content. Users could, for example, use buttons to pick up or drop a virtual item [22, 45, 85, 186, 206], or the scroll wheel to scale or zoom in/out the information [186, 218]. Blasko, Coriand, and Feiner also added a touchscreen on the projector's surface so that users could manipulate the displayed content with their fingers, such as zooming and panning the projected content [32].

Physical Manipulation of handheld SAR systems is commonly used as an interaction method. This approach is particularly prevalent in systems employing the flashlight or spotlight metaphor [32, 45, 186, 218, 247], which allows users to carry the unit and move in parallel with the projected surface to disclose information that was previously buried in the scene. Besides, as people wander around in the space, the projector can select and annotate virtual items with a cursor [22, 45, 85, 189, 249] or interact with virtual content created by other handheld systems [46, 248]. When users approach or step away from the surface, the proximity information can be leveraged to change the information granularity or zoom in/out of the projected content [45, 85, 193, 217, 218] to present different levels of details to the users. In addition to the translation movements, the orientation movement can also be used to interact with the virtual content. The wrist-worn projection system by Blasko, Coriand, and Feiner enables users to rotate their wrists to pan and zoom into areas of their interests [32] and the ClippingLight system [118] allowed users to tilt the system for zooming interaction and take the snapshots of both the physical and virtual content. The diverse combination of physical manipulations on the direct control SAR systems lead to a rich design space for people to engage with digital content [249].

Pens and smart devices are also used in coordination with the direct control SAR systems. A pen can be used to define interaction area [45] or annotate both the physical object and virtual content in combination with a system's physical manipulation [218]. Smart devices, on the other hand, offer multitouch inputs for target selection [120] or picture browsing [93]. They can also be placed over projected content, and show additional details within the screen [100, 208].

The desktop workspace is a hub of various interactions, necessitating the integration of multiple inputs to enhance its pervasiveness. Implementing mouse and direct touch is crucial because people are familiar with these interactions in the desktop space. Explicit and direct control of the system should also be incorporated. This drives the design of our system, as discussed in Chapter 3. Direct manipulation of the system creates a dynamic virtual display within the physical environment. The integration of various inputs enables varied interaction methods on an augmented desk with the dynamic peephole display, which is the focus of our work in Chapter 4. Results in Chapter 4 motivate our work in Chapter 5, which outline potential design considerations to the lamp's direct manipulation.

2.4 INTERACT WITH PHYSICAL ENVIRONMENTS VIA SAR

When an explicit and direct control is applied to a SAR system, it creates new interaction contexts with both the physical surfaces and objects.

2.4.1 *Dynamic Peephole Display on Physical Surfaces*

SAR transforms physical surfaces into interactive displays. While SAR with no user control creates static virtual displays, SAR with implicit and explicit user control generates dynamic virtual displays, revealing information only through this viewport on the augmented physical surfaces [42, 45, 249]. The interaction bringing virtual objects hidden in a larger virtual workspace into a smaller viewport display is often called *peephole interaction* [83, 84]. Butz and Kruger leveraged this idea to augment physical surfaces in a room dynamically, such as highlighting a book on the shelf or labelling an object with virtual information as needed [42]. While this spatial-aware interaction is commonly used in various environments [41, 195, 263], few studies explore how different factors, such as control mechanisms, peephole sizes, impact this interaction on augmented surfaces. Recent research by Kaufmann and Ahlström explored this interaction using a handheld projector on a wall, identifying target overshooting as a key characteristic of peephole pointing in augmented surfaces [120].

2.4.2 *Physical Objects Augmentation*

Physical Objects Augmentation refers to interactions where physical objects serve as input, output, or both. The augmented objects could be everyday objects like clocks, speakers and keyboards, or smart devices like smartphones and tablets. Early work, such as the Augmented Surfaces system [192], allowed users to annotate virtual content around physical objects and share information among them via the augmented surfaces. Building on this concept, Kane et al. developed the Bonfire system, which enabled photo transfer from smartphones to laptops using projected displays, demonstrating the potential for integration between digital and physical spaces [119]. As the technology advanced, handheld SAR systems evolved to become more versatile tools. With an explicit and direct control, PICOntrol [206] showcased how projectors could function as remotes to control physical objects like lights or television. People can also use these systems to enrich interactions on smartphones in the workspace, such as picture browsing [93], map navigation [208] and text inputting [100]. One aspect that has not been explored is to leverage the physical affordance of physical objects to support an explicit and direct manipulation of SAR and how these physical affordance can be leveraged to design interactions for augment surfaces and objects.

In Chapter 3, we leverage the physical properties of an architect lamp, and create an environment-aware SAR system on top it. Besides the lamp being augmented, the system also supports other physical objects augmentation, such as clock, mouse and smart devices within the augmented workspace. The dynamic peephole display enables new interactions and lead to unexplored human-factors and user behaviors questions that are related to the direct manipulation of an SAR system. Specifically, we address the coupled and decoupled control target acquisition for the dynamic peephole interactions on the augmented surfaces in Chapter 4.

2.5 SUMMARY

This chapter summarizes the literature on pervasive desktop computing from a technical perspective. Reviewing prior methods and systems helps us to identify challenges and design considerations to address research objective.

Prior methods all focus on a specific problem for pervasive desktop computing or fail to establish a bidirectional connection that allow virtual content to move across the Environment dimension. To address these limitations, we build an augmented SAR lamp system described in Chapter 3. It enables people to engage with digital information with methods they are familiar with in the workspace, but also allow them to augment the physical surfaces, objects and devices.

As users need to manipulate the lamp to transform the physical surfaces into interactive displays, it creates a dynamic peephole display on the augmented surfaces. Since one hand is dominated by lamp manipulation, we investigate the impacts of control mechanism on target acquisition for the dynamic peephole interactions on the augmented surfaces in Chapter 4.

Results in Chapter 4 reveal distinct interaction patterns in the same search phase, and different interact strategies overall for two control mechanisms. We extend this observation and investigate a more general form by investigating effects of what users intend to do with a virtual target impacts how they plan and perform the initial target acquisition in Chapter 5.

3 LUXAR: A DIRECT MANIPULATION PROJECTED DISPLAY TO EXTEND AND AUGMENT DESKTOP COMPUTING

In this chapter, we describe the design and implementation of an explicit and direct manipulation SAR system LuxAR, by leveraging the unique physical characteristics of an architect lamp. This augmented lamp establishes a bidirectional connection across the *Environment* dimension and maintain *Awareness* of the physical environment and users' activities.

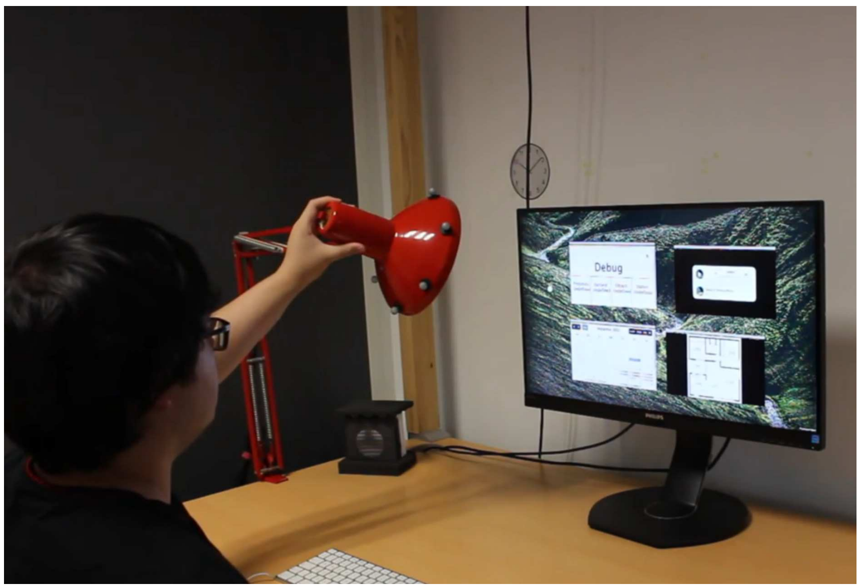


Figure 3.1: Users use LuxAR to move desktop windows onto physical surfaces.

To identify our design direction, we conducted an informal exploration of the desktop workspace from nine people (see Figure 3.2). Our initial motivation was to examine how people configured their desktop workspace, what devices and displays are used, and what can be implemented to augment the existing desktop environment to enrich the information access and address people' diverse interaction needs.

People often rely on multiple displays and devices to create an expanded desktop space for multitasking and information coordination [27, 75, 105, 264]. However, physical space constraints, optimal device placements, and cost often pose challenges [105, 264]. A mixed reality (MR) head-mounted display (HMD) can be used to expand desktop computing space [137, 172]. However, using an HMD creates an isolated digital workspace [172] and can induce fatigue [28].

Our observations and prior work led us to think about leveraging available physical surfaces and augment objects to enhance user experiences, and

Figure 3.2: Sample images of desktops collected for initial observations.

Spatial Augmented Reality (SAR) is one of the solutions, which uses one or more projector-camera units (pro-cams) for projection mapping on physical surfaces to transform them into interactive displays [115, 190].

SAR has been used to extend desktop computing to surfaces [130, 192, 241], but in a limited way with a single fixed pro-cam. Some SAR systems use multiple fixed-position pro-cams to cover more surfaces [115, 190, 192] for presenting information. However, projection space cannot be adjusted once the system is running. One way to overcome this limitation is to use steerable pro-cams [116, 251], but users cannot explicitly control the projection space. Handheld pro-cam systems supporting explicit and direct input have been proposed [22, 45, 104, 249], but these must be held in the air for aiming and input movements, fully occupying at least one hand and leading to fatigue. To support explicit and direct manipulation with a pro-cam, a method to hold the pro-cam in space would be ideal, especially to support long-duration tasks such as placing additional desktop information on surfaces.

Figure 3.3 demonstrated our process to build our prototypes and find such an ideal method. In the first prototype, we installed a mini portable LED projector and a camera installed on a manipulable tripod arm, and built the system in C ++ with projection mapping method for calibration and yolo4 [34] for object registration. However, we soon noticed that the tripod arm required considerable physical efforts to manipulate and was unstable during manipulation. Revisiting our collected workspace (Figure Figure 3.2), we found an ideal replacement to achieve our goal —- a lamp.

We then developed our second prototype using a lamp, but we encountered new challenges. For the LED projector, we needed manual adjustment to

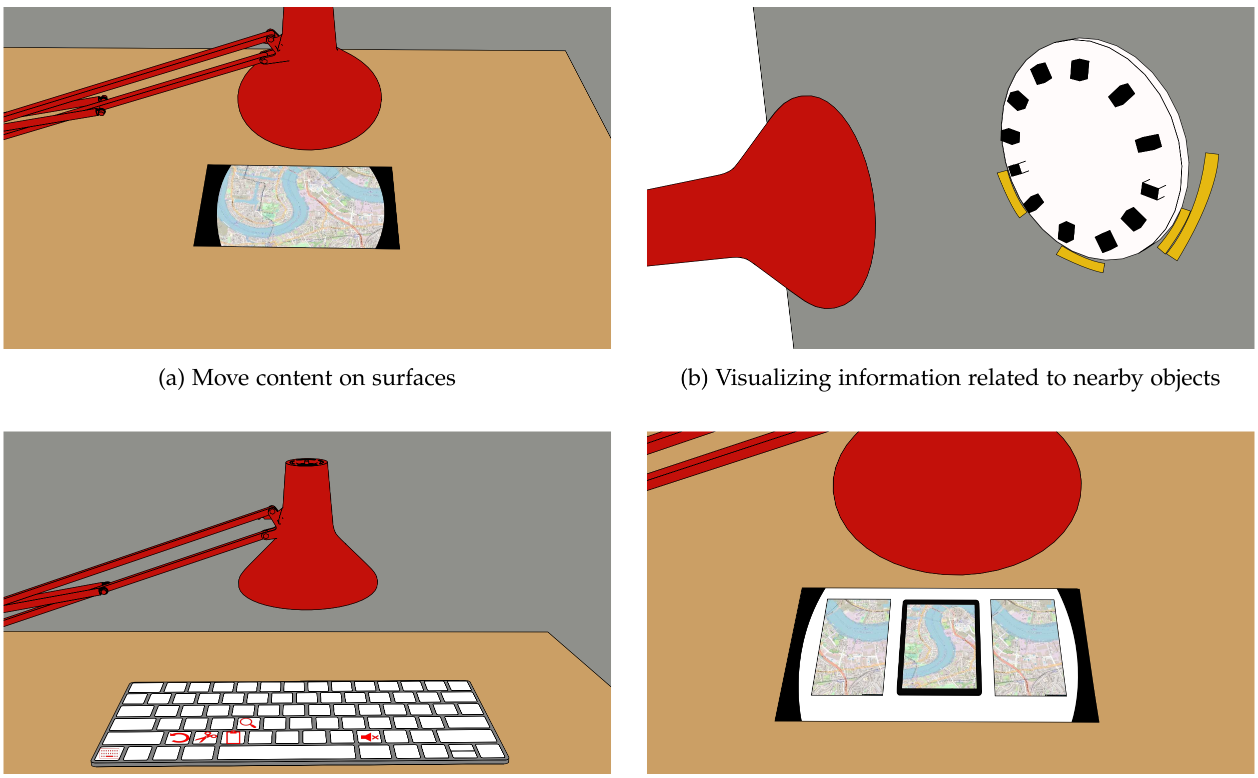
Figure 3.3: Earlier prototypes of our system by iteration order.

achieve the right focus, and the projection mapping method struggled with the dynamic repositioning of the pro-cam unit. To address these issues, we replaced the LED projector with a laser projector that supported autofocus and used reflective markers and the OptiTrack system for objects' registration and tracking, similar to previous handheld projector systems [45, 46]. Thus, in our third prototype, we considered the laser projector as the "light source" for a lamp, mounting it on a manipulable monitor arm for improved stability and control. Meanwhile, we discarded previous code-base and rebuilt the whole system from scratch in Unity and *C#*. However, we encountered an unexpected challenge: the monitor arm's strings were too taut, making it difficult to maintain desired positions. This led us to our final design choice — an architect lamp. This solution struck an ideal balance between ease of manipulation and stability, providing the well-behaved mechanical properties we needed for our system.

An architect lamp has articulated, counterbalanced arms, that can support a pro-cam in space. This means the lamp can be aimed at a surface, and it will maintain its pose after manipulation. The Lantern demo [135] and the LuminAR prototype [139] both integrate a pro-cam in an architect lamp and leverage these adjustment and support properties. However, the Lantern requires a phone for all input relegating the role of the lamp to be only an adjustable display. LuminAR actuates the lamp using motors, with display position controlled indirectly through gestures performed in front of the lamp with touch input to interact with content. This adds significant complexity and cost. Importantly, neither uses an augmented lamp as a form of direct manipulation on its own.

We use a pro-cam mounted in a standard architect lamp for explicit and direct manipulation (Figure 3.1) of a SAR interaction space to extend and augment desktop computing (Figure 3.4). Our new design space, with proof-of-concept LuxAR system, demonstrates a novel form of direct manipulation leveraging unique characteristics of physical lamp movement. With a single button (mounted on the lamp hood like a standard switch), we show how this can be used to reposition displays, interact with applications, and adapt content to various surfaces, devices, and objects. To maintain the flow of desktop computing, the approach can integrate standard input from the mouse, as well as mobile devices with touch and pen.

Our work makes three main contributions: (1) a new interaction design space for physical direct manipulation using an articulated pro-cam lamp; (2) the LuxAR proof-of-concept system demonstrating usage applications in various scenarios; and (3) results of a two-phase user study evaluating the potential of such prototype and design implications for explicit and direct manipulation systems to extend and augment desktop computing. Together, these contributions establish a promising way to connect current desktop computing with the surrounding physical desktop space.



(a) Move content on surfaces

(b) Visualizing information related to nearby objects

(c) Augmenting keyboard with shortcut keys

(d) Extending interaction space of mobile devices

Figure 3.4: LuxAR is an architect lamp instrumented with a projector to extend direct manipulation interfaces into the physical surroundings.

3.1 PROJECT-SPECIFIC BACKGROUND

In Chapter 2, we discussed how SAR projection systems desktops for extending desktop computing and conclude with three user control mechanisms of the projected virtual display. We leverage the conclusion and present past examples of lamps as input or output devices. Table 3.1 summarizes our work compared to previous systems.

		User Control			Functions				Surfaces						Inputs							Modality		
	Form	Implicit	Explicit Indirect	Explicit Direct	Display	Reposition	Interact	Track	Wall	Table	Ceiling	Objects	Devices	Env. Awareness	Handheld	Mouse	Pen	Touch	Device	Gestures	Voice	Attached	Detached	User Eval
Our Work	**Architect**			✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓	✓
Lantern Demo [135]	**Architect**			✓	✓	✓			✓	✓	✓	✓			✓				✓			✓	✓	
LuminAR [139]	**Architect**		✓		✓	✓			✓	✓								✓		✓			✓	
MotionBeam [249]	Torch			✓	✓	✓	✓		✓					✓	✓							✓		✓
Cao et al. [45]	Torch			✓	✓	✓	✓		✓	✓		✓			✓		✓					✓	✓	✓
Beardsley et al.[22]	Torch			✓	✓	✓	✓		✓			✓			✓							✓		
PICOntrol [206]	Torch			✓	✓	✓	✓		✓			✓		✓	✓							✓		
SideBySide [248]	Torch			✓	✓	✓	✓		✓						✓							✓		
AR Magic Lantern [104]	Torch			✓	✓	✓			✓			✓		✓	✓							✓		
Omnilantern [260]	Lantern			✓	✓	✓			✓	✓		✓			✓							✓		
HideOut [247]	Torch			✓	✓				✓	✓		✓			✓							✓		
Project LFX [182]	Ceiling	✓			✓	✓			✓	✓		✓		✓					✓		✓		✓	
HuddleLamp [184]	**Architect**							✓					✓						✓				✓	
IllumiShare [117]	**Architect**				✓					✓							✓	✓					✓	✓
AR Lamp [124]	**Architect**				✓					✓							✓	✓					✓	✓
FACT [138]	Tabletop				✓					✓							✓						✓	
Xiao et al. [259]	Ceiling				✓					✓								✓					✓	
Lamposcope [239]	**Architect**				✓					✓							✓						✓	

Table 3.1: Previous interactive lamp systems compared to our work (sorted by relevance, see text for comparison criteria).

3.1.1 *Lamps as Interactive Systems*

Lamps are ubiquitous objects in many interactive systems and offer various opportunities for interactions based on their forms. When not manipulated, they serve as strong supports to hold a system and allow users to interact with the content using different inputs. Using only a camera mounted inside a lamp, HuddleLamp [184] tracks devices and recognizes configurations, so users can annotate the same information across different devices. But this is only a computer vision input platform. Some systems use a fixed display, for example integrating a pro-cam into a ceiling light [259] or a stationary lamp on the desk [117, 124, 138, 239]. These systems do not focus on input, limiting interactions to specific regions with hands or a pen. The form factor of a torch (i.e. a "flashlight") [22, 45, 104, 247–249] and lanterns [260] have inspired handheld display devices for users to interact with virtual information directly. Of course the handheld nature means at least one hand is occupied during active usage to maintain the position and to perform direct manipulation for input. This makes it challenging to establish sustainable augmented physical environments for desktop computing spaces.

In contrast, articulated lamps, especially architect lamps, featuring unique mechanical structures, provide flexible manipulation and maintain their pose after manipulation. This creates an opportunity for users to interact with information using interactions and inputs that are *attached* or *detached* to the direct manipulation of an architect lamp. In addition to hold a pro-cam unit in space [117, 124, 239], for example, the Lantern demo [135] allows users to rotate the lamp's head to adjust projected displays on a desired surface, and then use phones to change the content. LuminAR [139] integrates architect lamps with robotic systems and enables users to control the projected display through midair gestures, and interact with content through direct touch. However, unlike torch or lantern inspired systems, both fail to consider designing interactions based on their direct manipulation and poses, and are unaware of changes of physical spaces to adapt content accordingly. Critically, while various types of lamps have been proposed as interactive systems, there are surprisingly few user evaluations of these kinds of systems.

To the best of our knowledge, the direct manipulation potential of architect lamps has not been investigated yet in order to interact with virtual content and extend and augment desktop computing. We examine potential design possibilities in this work and introduce LuxAR as a system to elevate user experiences from the desktop space to various physical surfaces.

3.2 LUXAR DESIGN GOALS

We want to fully leverage architect lamp properties for direct manipulation and ability to remain stationary after control. The goal is to design an explicit and direct SAR system to extend and augment a desktop computing space onto nearby physical environments with an awareness of the immediate environment. In contrast to most torch-based systems using direct manipulation input, it should also support interaction with content using other inputs. We summarize with the following designs goals:

- DG1 Content interactions can be attached to lamp's direct manipulation, like rotating for repositioning or adjusting distance to change displayed information.
- DG2 When adjusting the lamp, content can appear on various desktop surfaces like tables, nearby walls, and the ceiling. Content could also be projected onto physical objects and mobile devices on these surfaces. The floor is excluded due to clutter and user movement.
- DG3 The system should detect changes in physical surfaces, objects, and devices and adapt the displayed content accordingly.
- DG4 Content interactions can also be detached to the manipulation of the lamp, allowing users to use other inputs on the desktop.

To address these design goals, we explored potential application scenarios through physical papers and mock-up interactions (Figure 3.5). This parallel process allowed us to visualize and simulate how content could be dynamically adapted based on the physical manipulation of the architect lamp, physical surfaces, objects and devices where the content is positioned. By simulating different scenarios, we identified how these design goals can be leveraged for proposing interaction techniques.

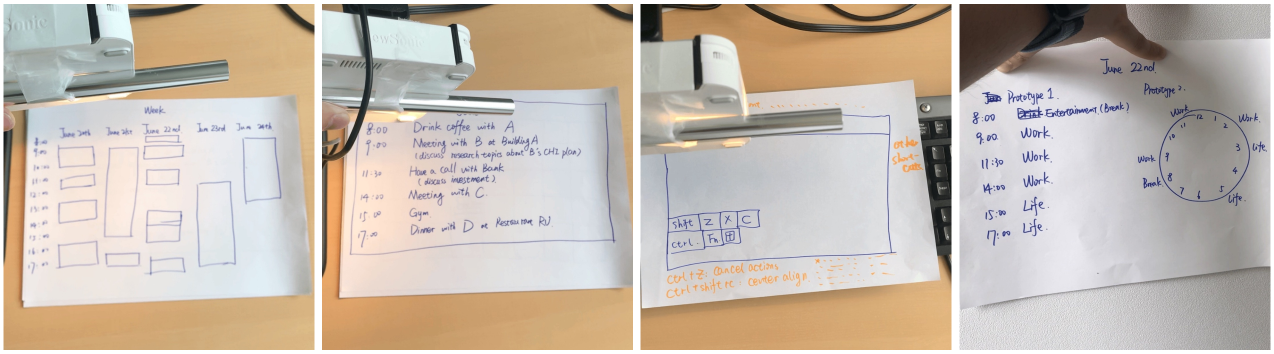


Figure 3.5: Mock-up interactions of the calendar scenario for LuxAR with physical paper and a table light.

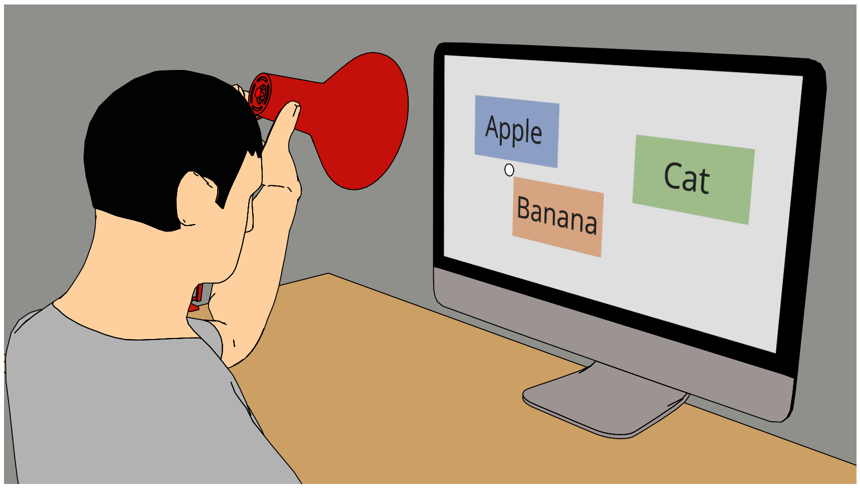
By implementing these design goals, our approach serves as both an input and output device to connect desktop computing with the nearby desktop environment through explicit and direct manipulation. A key aspect is how it leverages physical characteristics of an architect lamp, occupying a unique position relative to previous work (Table 3.1).

3.2.1 *Content Repositioning [DG1, DG2]*

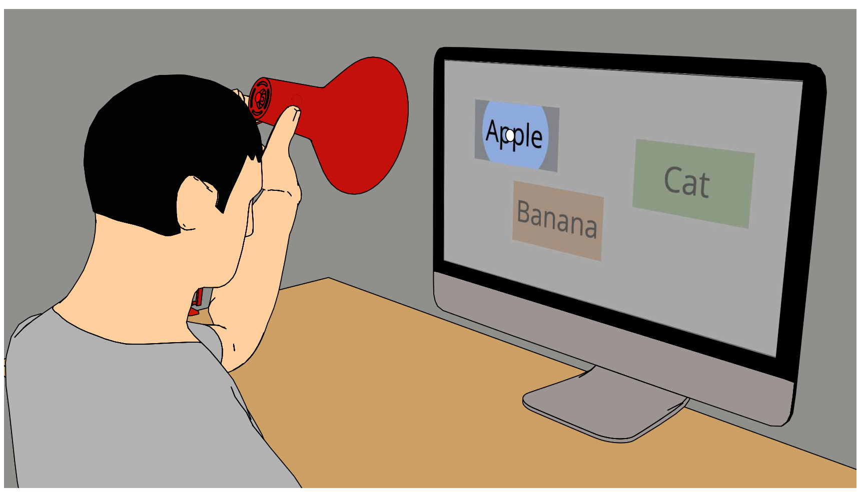
Content can be moved beyond monitor boundaries onto surrounding surfaces such as desks, walls, and ceilings. Ray-casting is used to interact with virtual content [22, 32, 45]. The lamp head is the ray origin with the lamp head angle determines the ray direction. A button on the lamp shade can be pressed with an index finger or thumb while manipulating the lamp. A spotlight metaphor visualization [186] highlights the location where the lamp points. Two variations of this spotlight visualization are used.

When pointing at a monitor, the visualization of the ray intersection point is a small white dot, essentially forming a "lamp cursor" (Figure 3.6a). To emphasize the dot, the rest of the screen dims. Hovering the lamp cursor over an application window further dims the screen and highlights the window in an oval shape (Figure 3.6b), indicating it is select-able by the lamp. To move the window to a surface outside the monitor, the user presses and holds the lamp button to drag, much like dragging a window with a mouse (Figure 3.6c). Releasing the button drops the window onto the surface. The window remains anchored in place until it is picked up again.

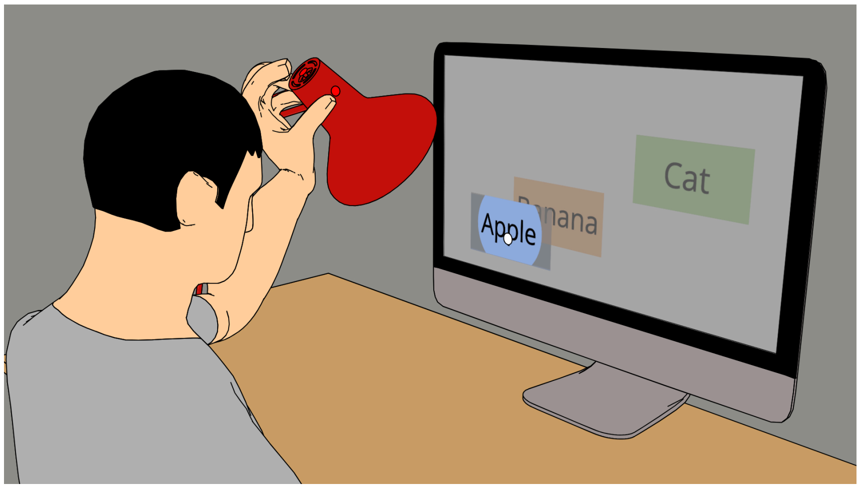
When pointing at surfaces outside the monitor, the spotlight visualization changes according to four modes based on the visibility of the window and the interaction status (Figure 3.7). In radar mode, a circular display appears when the lamp is not directed at any window. Colored icons within indicate the direction and location of an anchored but not visible window (Figure 3.6d). Hover mode activates when the lamp hovers above a virtual window, displaying a large oval shape for detailed examination (Figure 3.6e). Clicking or pressing a button on the window transitions to focus mode, allowing the window to receive additional input events. Clicking on an empty



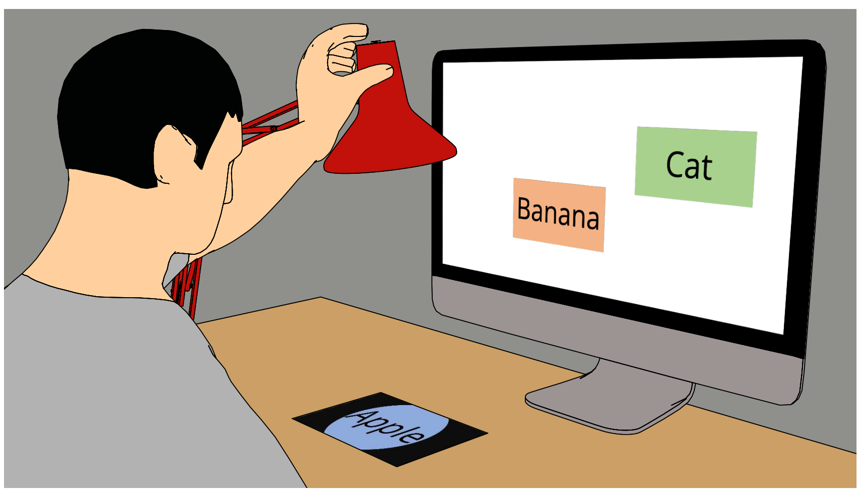
(a) Point at the monitor.



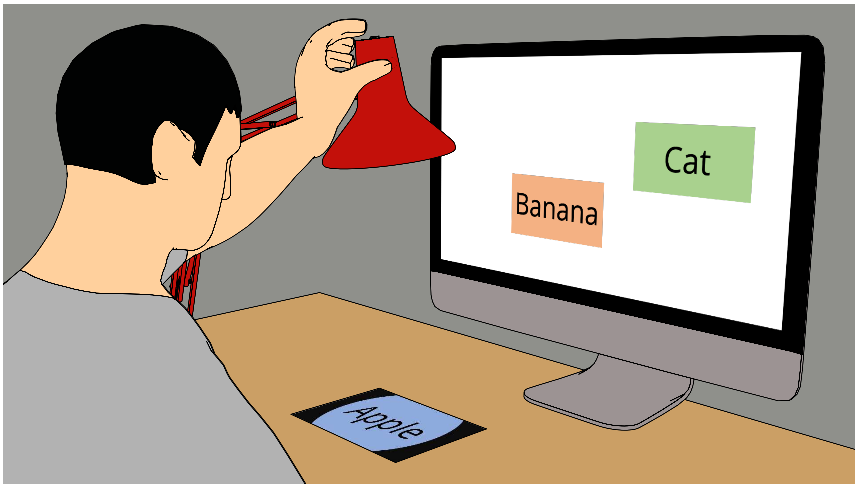
(b) Hover on the monitor.



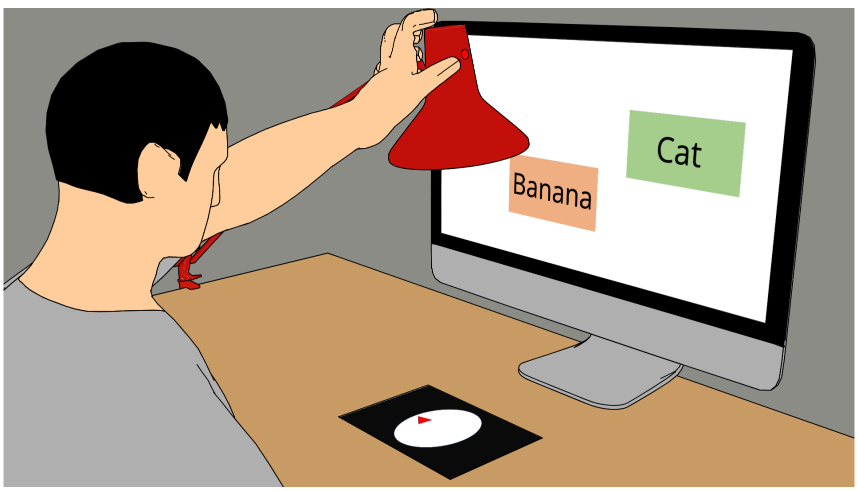
(c) Drag on the monitor.



(d) Drag on the surface.



(e) Focus on the surface.



(f) Radar on the surface.

Figure 3.6: From (a) to (e), users manipulate the LuxAR to move an application from the monitor onto a nearby surface.

area reverts to hover mode, and moving the lamp away from the window exits hover or focus mode, returning to radar mode. Users press and hold the button to enter drag mode. The lamp shows a smaller oval display to maintain contextual visibility during dragging (Figure 3.6f). The window maintains orientation invariant and aligns parallel to the user's seated location.

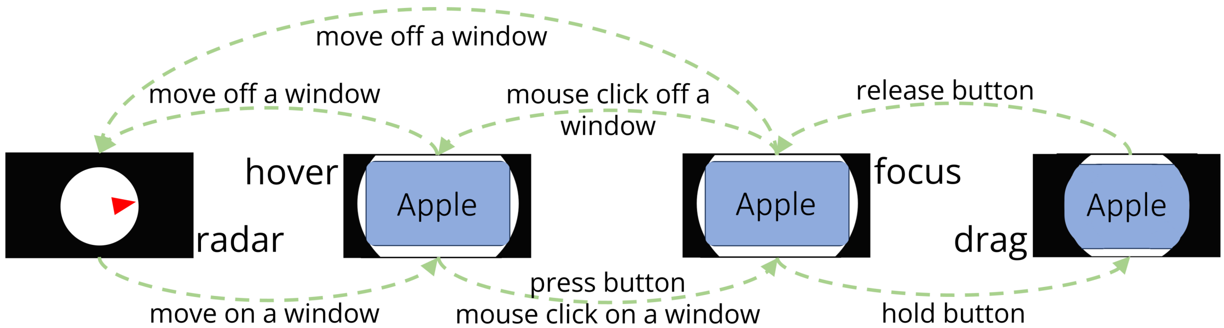


Figure 3.7: Interaction state machine with lamp display styles.

3.2.2 *Surface Adaptation [DG1, DG2, DG3]*

When a user moves a virtual window across different surfaces, like from the desktop to a desk, walls, and the ceiling, both the lamp display size and window visibility change, impacting user behaviors around the desk: A window on the desk is easily visible to a single user, while on the wall, it becomes noticeable to multiple users; users may find it less convenient to tilt their heads upward when seated to view information on the ceiling; yet, when users move away from the desk, information displayed on the ceiling serves as an ambient cue, easily noticeable from a distance.

We employ design principles inspired by Vogel and Balakrishnan's work [234] to facilitate interaction transitions. Our approach allows explicit manipulation of the lamp to reposition content on different surfaces, adapting the virtual window to distinct interaction spaces: the desk for personal interactions, the wall for public interactions, and the ceiling for ambient interactions, as shown in Figure 3.8. With all the information located on the desk, this space serves as a decision-making platform for users to determine the optimal placement and presentation of the virtual window based on their preferences. When the window moves to the wall, it becomes visible to a larger audience, fostering collective information sharing. On the ceiling, it may escape the seated user's immediate notice but serves as an ambient cue for others.

3.2.3 *Transition Granularity [DG1, DG2]*

Building on prior work that used proximity to adjust the detail and granularity of information in virtual content, such as PenLight [217], PaperLens [222], the revealing flashlight [193] and Cao and Balakrishnan's handheld system [45],

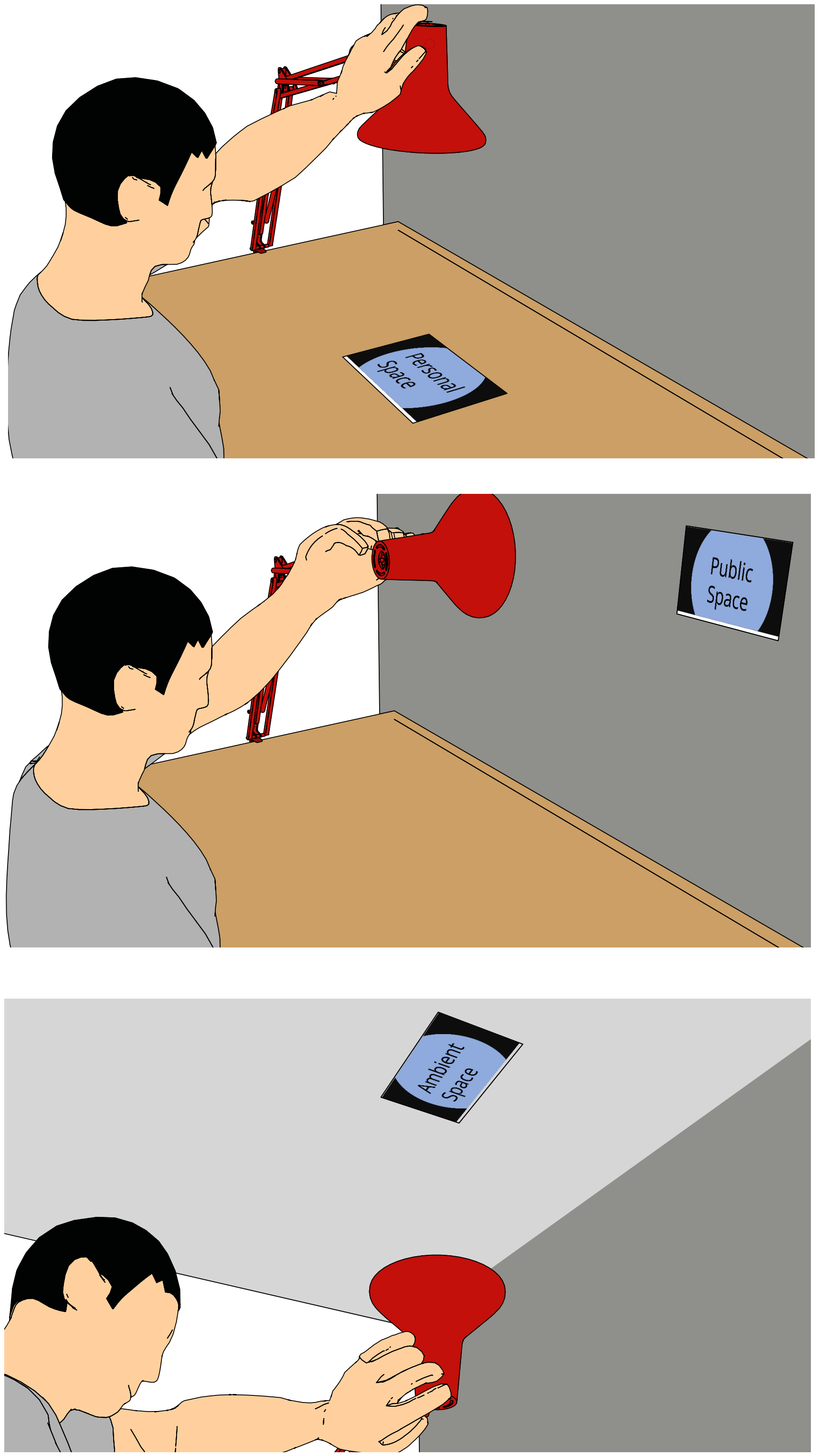


Figure 3.8: The content in the window adapts to the change of surface when users manipulate LuxAR.

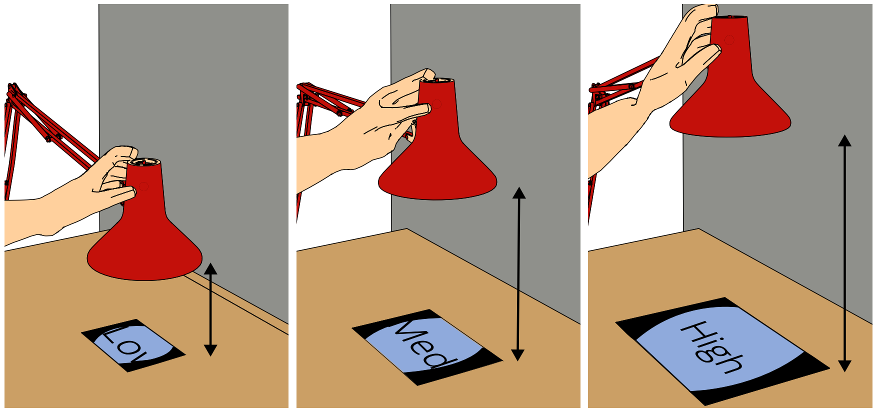


Figure 3.9: Adjusting the height alters the context of the window to accommodate the changing size of the lamp display.

we embrace a comparable approach. Proximity, in our context, is defined as the distance between the lamp and the projected surface, categorized into three levels: High, Medium, and Low. When our lamp is oriented to the desk surface, its height can be adjusted between approximately 7 cm and 80 cm, maintaining a stable position. Given that the size of the lamp's projected display is influenced by its height adjustment, ensuring optimal visibility and interactivity of the projected content within the lamp display becomes essential. Based on our lamp's physical shape, we empirically established threshold values (25 cm and 50 cm) to determine transitions between different proximity levels for the desk surface. We adopted these same values for the wall for simplicity, and they are ignored for the ceiling because of the absence of a manipulation axis.

At each proximity level, as the lamp's distance changes, the content and size of the virtual window remain consistent. This enables the system to simulate a zooming effect by dynamically adjusting the display size of the lamp. When the height of the lamp crosses a proximity threshold and transitions to a different level, the targeted virtual window adjusts its content, similar to the PaperLens technique [222]. Specifically, when the lamp is raised and positioned farther from the surface, the size of the lamp display increases. This exhibits higher information density in the virtual window, providing a broader context. Conversely, as the lamp is gradually brought closer to the surface, the size of the lamp display decreases. The virtual window reduces information density while enhancing information quality, offering a focused view of the virtual window.

3.2.4 *Object and Device Augmentation [DG2, DG3]*

The desk space, with various displays, devices (*e.g.* smartphones), and tangible objects (*e.g.* keyboards, mice, and clocks), requires thoughtful consideration for the system's awareness of its surroundings. Taking into account the timing

(Temporary vs. Permanent) and presentation (Overlay vs. View) aspects of augmentations to physical objects and devices, we suggest three interactions: Temporary Overlay, Permanent Overlay, and Temporary View. We exclude Permanent View since LuxAR inherently transforms physical surfaces into permanent displays for presenting information.

Permanent Overlay (PO) persistently augments physical objects, anchored to their location, revealed by the lamp hovering. For example, a keyboard is always augmented and the shortcuts for an application are overlaid on it when hovering over it with the lamp.

In contrast, Temporary Overlay (TO) consists in temporarily enhancing a physical object based on the content displayed in a virtual window. A user picks up and drops a virtual window on a physical object. The window disappears and the content blends into the object, aligning with its context and shape. To retrieve information from the object, users replicate the same action on the object, transforming the blended content back into a window. An example is to augment a clock with virtual window information, such as reminders, calendars, and weather forecasts, allowing users to change the augmentation by dropping different windows.

For mobile devices on the desk, Temporary View (TV) enables users to expand the small display of the device using the lamp. The procedure is akin to Temporary Overlay: users pick up and drop a virtual window into the mobile device. Subsequently, the virtual window is transferred to the device, and additional views are displayed around it to enrich interactions.

3.2.5 *Other Inputs [DG4]*

Device Interaction

Temporary View enables users to interact with projected virtual content via touch screens on mobile devices, the modified content can then synchronize across various devices, surfaces, and desktops.

Mouse Interaction

Given LuxAR's expansion of existing desktop environments into physical spaces, maintaining mouse-based interaction is crucial, as they have demonstrated efficiency in SAR [101, 123]. The virtual cursor is present on both the monitor and the lamp display, seamlessly transitioning between them. Users can edit content on physical surfaces using the mouse when the cursor is in the virtual window. Additionally, the spatial relationship between the lamp and the physical display allows users to reposition the virtual cursor. This design also extends cursor functionality to augmented mobile devices, enabling interaction through both touch input and mouse control with Temporary View augmentation.

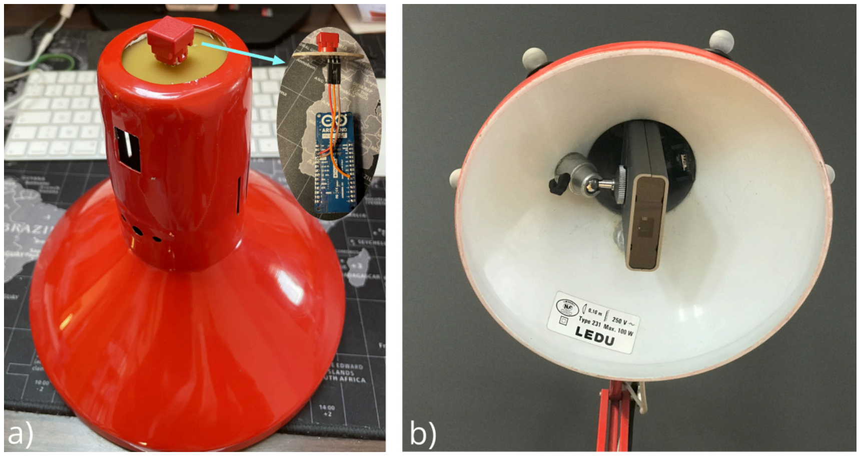


Figure 3.10: System hardware: (a) the standard lamp switch is replaced with an input button (b) a pico laser projector mounted inside the shade, and motion tracking markers attached on the lower bezel of the lamp shade

Pen and Touch Interactions

We enable touch interaction through a ring-mounted marker on the index finger and pen interaction using a registered pen, both tracked by the OptiTrack.

3.3 proof-of-concept system

3.3.1 *Augmented Architect Lamp*

Our proof-of-concept prototype was built in two simple steps. First, we created the LuxAR prototype by modifying a Ledu architect lamp. We integrated a Nebra AnyBeam laser projector[1] beneath the lamp, measuring 103 mm (Length) × 50 mm (Depth) × 19 mm (Height). This projector supports autofocus and provides a resolution of 1280 × 720p at 60 FPS. Additionally, to enable the input capability, we installed a button on top of the lampshade, using an Arduino MKR Wi-Fi 1010 (Figure 3.10). This design allows users to use a finger, typically their index or thumb, to press the button while the other four fingers hold the lamp firmly. Next, we installed six reflective markers around the lower bezel of the lamp shade, which were tracked by an OptiTrack system[2]. These markers formed a rigid body used to update the position and orientation of the lamp in the system. Further details on environment reconstruction and object registration and tracking are provided below.

1 https://github.com/NebraLtd/AnyBeam
2 https://optitrack.com/

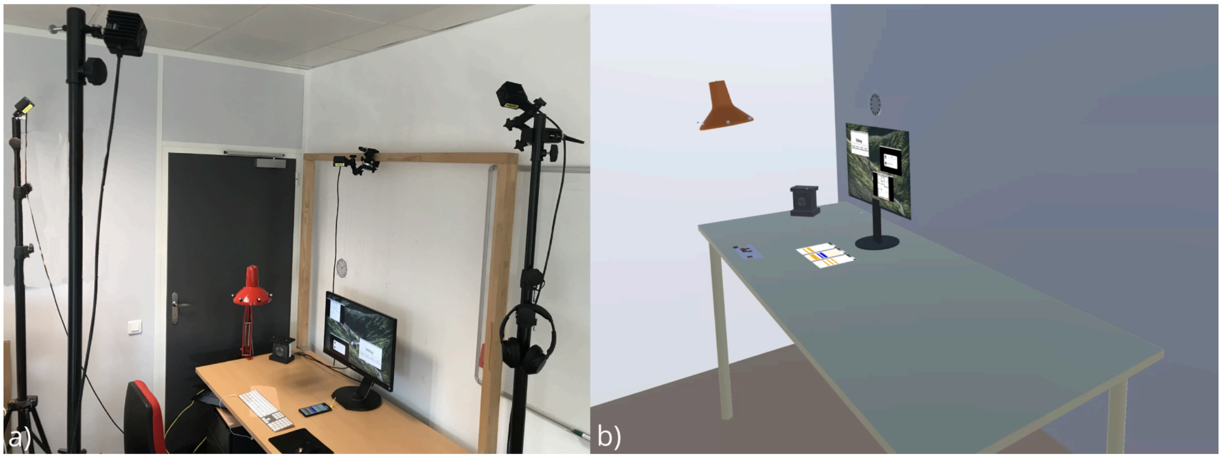


Figure 3.11: (a) desk space with four OptiTrack cameras and physical objects and devices. (b) reconstructed space in Unity.

3.3.2 *Environment Setup and Object Registration*

We employed the OptiTrack system, featuring four cameras, to track and partially reconstruct the desktop workspace (Figure 3.11). Markers were placed on the wall, desk, and monitor to define their location and associate virtual objects. We built a custom pen equipped with three markers (see Figure 3.16) and a unique OptiTrack ID to track its position and orientation. The tip of the pen was then used to register the pose and dimensions of static objects (*e.g.* a desk keyboard) by placing the tip at the outline of them and marking its positions. For objects with mobility, such as the lamp, smartphone, and ring, markers were attached and assigned unique IDs if possible, allowing continuous tracking of their position and orientation.

3.3.3 *Touch Tracking and Detection*

There is a challenge with the single marker on the fingertip, as it receives a different ID each time the cameras lose tracking due to occlusion or when dummy markers appear. To mitigate this issue, we implement a frame-by-frame exclusion of tracked points registered for moving objects (the lamp, the phone, and the pen). Then, when the ID for the index finger marker is missing, we re-assign that ID to the closest individual marker that is below a 10 cm range of the previously known position of the index. If no point is found, we keep the previously known position of the index.

To accommodate different thicknesses and poses of the index finger, we perform a calibration phase for each participant. Users place their finger in contact with the desk to simulate a touch interaction and keep that pose for at least 5 seconds. The average distance between the marker and the desk surface is measured during that time to define a threshold distance used to consider whether there is a touch with the surface.

3.3.4 *Implementation*

Our system was developed in Unity 2021.3.8f1, using Motive software 1.10.2 and the corresponding OptiTrack Unity Plugin 1.4.0. Camera settings in Motive included a frame rate of 180, exposure of 40, brightness threshold of 254, and LED illumination of 15. Cameras used low-range gain, infrared spectrum filter, and precision mode. While the physical lamp's position and orientation were tracked, the laser projector within remained hidden, with its exact orientation and position unknown. For consistency between virtual and physical environments, the laser projector and virtual camera were treated as identical. We fine-tuned the virtual camera's sensor sizes and offsets within Unity, aligning the calibration using the CS-200 calibration square of the OptiTrack system. This manual calibration process ended when the virtual and physical markers were approximately aligned in position, size, and orientation.

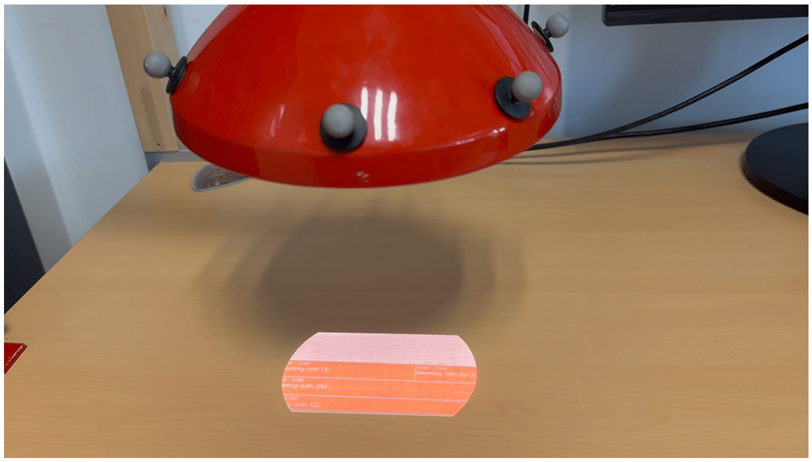
Our software, a virtual desktop that manages application windows on various surfaces, featured independent WebViews for each virtual window. These were supported by the Embedded Browser package[3]. Windows adapted their content based on factors such as current location (*e.g.* the monitor, desk, wall or ceiling), awareness of the physical environment (*e.g.* clock, speaker, phone), lamp height, and user input (*e.g.* lamp, mouse, pen, touch). Applications were implemented outside of Unity and hosted on a local server. Virtual window content was displayed based on the URLs provided and responded to events triggered by user input, and communication between the lamp system and the phone, including touch events, was established over a local network.

3 https://assetstore.unity.com/packages/tools/gui/embedded-browser-55459

3.4 USAGE SCENARIOS & APPLICATIONS

To demonstrate the supported interactions, we provide potential usage scenarios with three applications: a calendar, a music player, and an architectural design drawing tool. The accompanying video [4] provides full demonstrations.

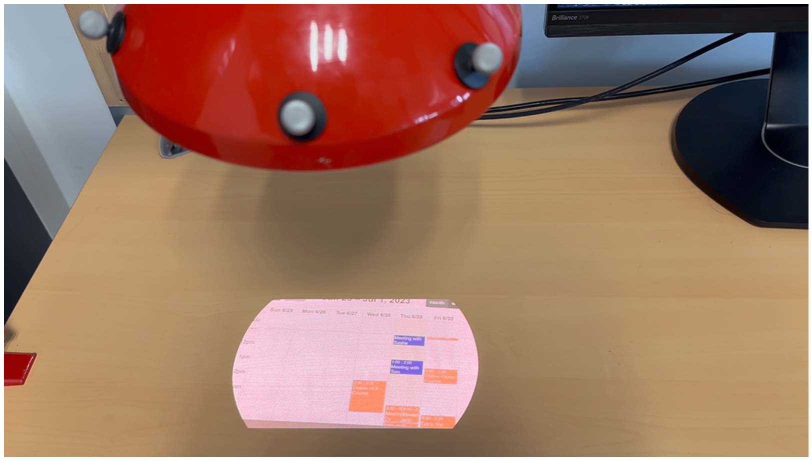
Calendar



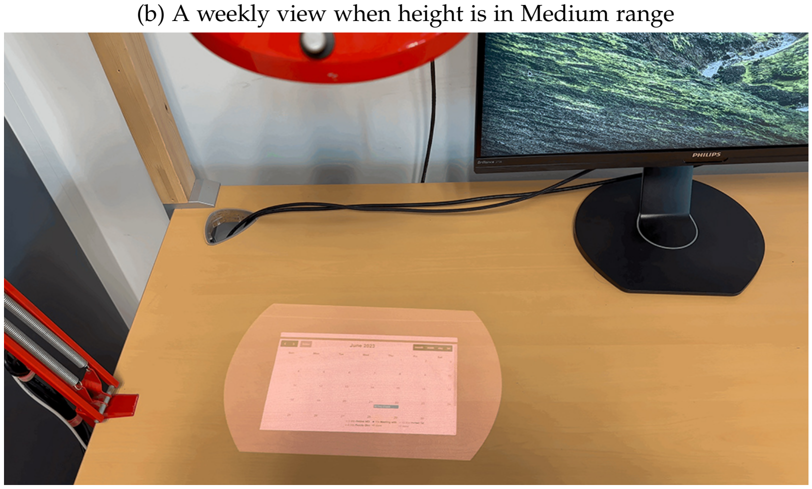
(a) A daily view when height is Low range

Alice uses LuxAR to streamline her calendar management. By pointing at the calendar window on the monitor, she can easily drag it out, placing it on her desk with a press-and-hold button action. The height adjustment feature of the lamp allows Alice to access various levels of detail in the calendar window. At a medium height, she views the weekly schedule (Figure 3.12b). Bringing the lamp closer reveals the daily schedule (Figure 3.12a), while raising it higher displays the monthly view (Figure 3.12c). Alice moves the calendar window from the table (Figure 3.13a) onto the wall (Figure 3.13b), showcasing schedules to her colleagues. Alternatively, placing the window on a clock creates a visual timeline (Figure 3.14a). During breaks, she moves it to the ceiling, turning it into a countdown timer for the next event, reminding others and herself (Figure 3.13c). Alice can edit the calendar events on the lamp display using her mouse or finger. Hovering the lamp over the keyboard reveals the calendar application shortcuts (Figure 3.14b). She can also transfer the calendar to a nearby mobile device, to better manage the events on her phone (Figure 3.14c).

4 https://youtu.be/dTXoue6qoVw



(b) A weekly view when height is in Medium range

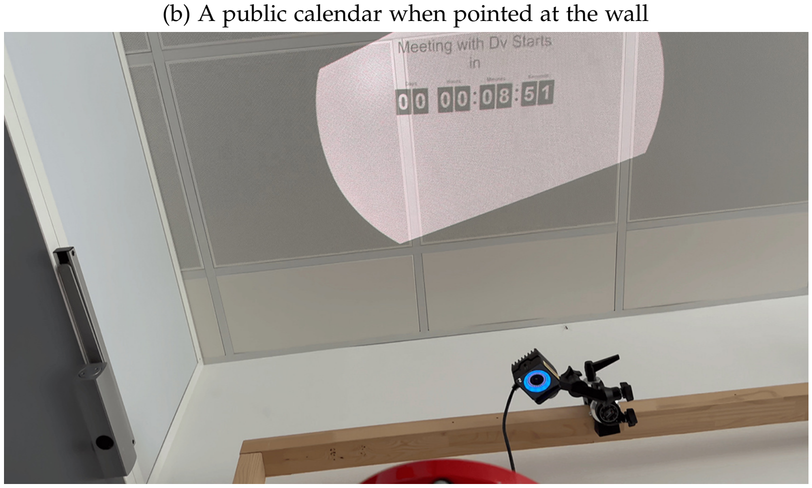


(c) A monthly view when height is in High range

Figure 3.12: Pointing the lamp at the content with different heights can change how content is presented, for example in a calendar application.

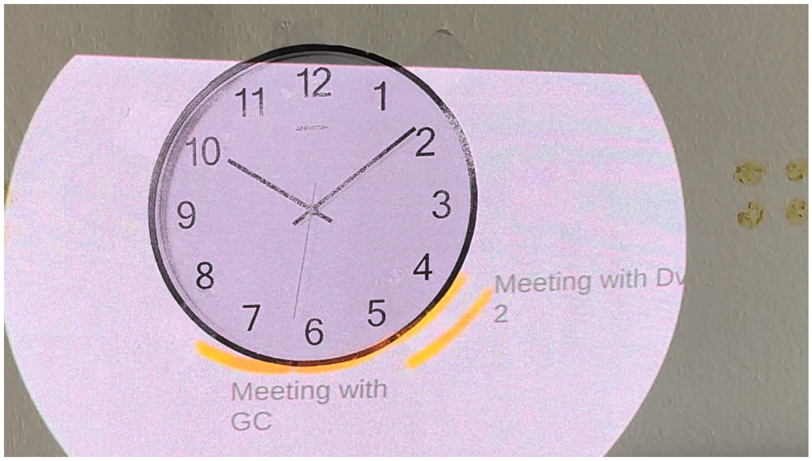
(a) A private and public calendar when pointed at the desk

(b) A public calendar when pointed at the wall

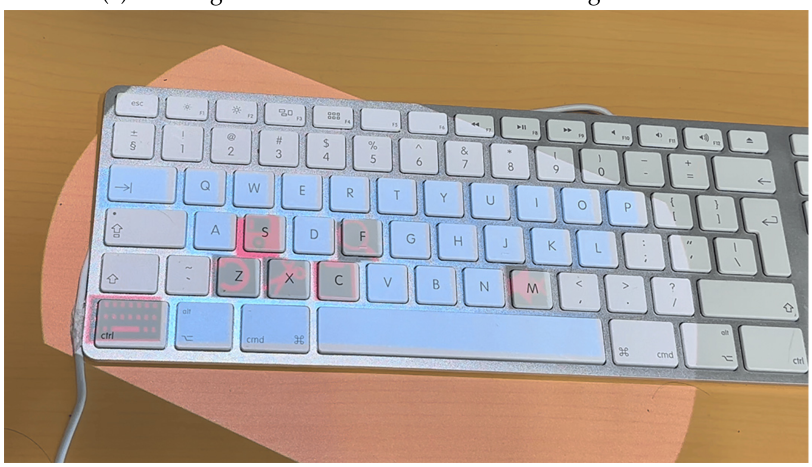


(c) A countdown to the next event when pointed at the ceiling

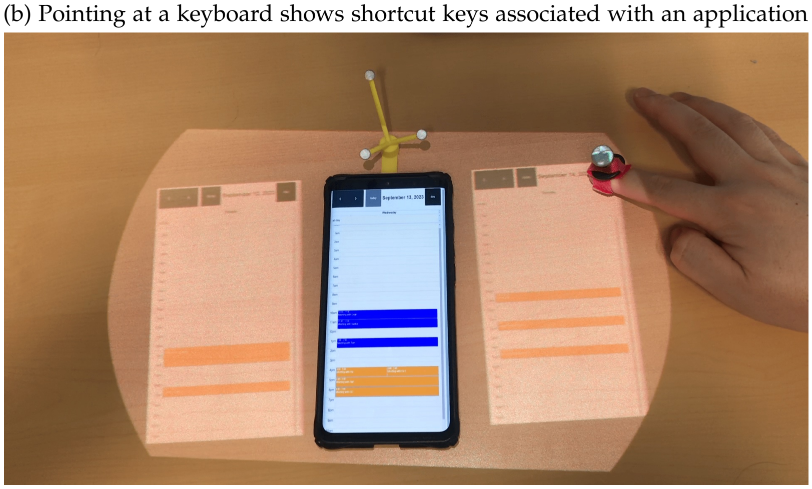
Figure 3.13: Pointing the lamp at different physical surfaces can change how content is presented, for example in a calendar application.



(a) Pointing at a clock shows scheduled meetings next to it



(b) Pointing at a keyboard shows shortcut keys associated with an application



(c) Pointing at a mobile phone with the calendar application shows the previous and following day next to the phone

Figure 3.14: Pointing the lamp at specific objects or devices can augment them with content.

Music Player

(a) Lamp at medium height shows a list of songs with covers

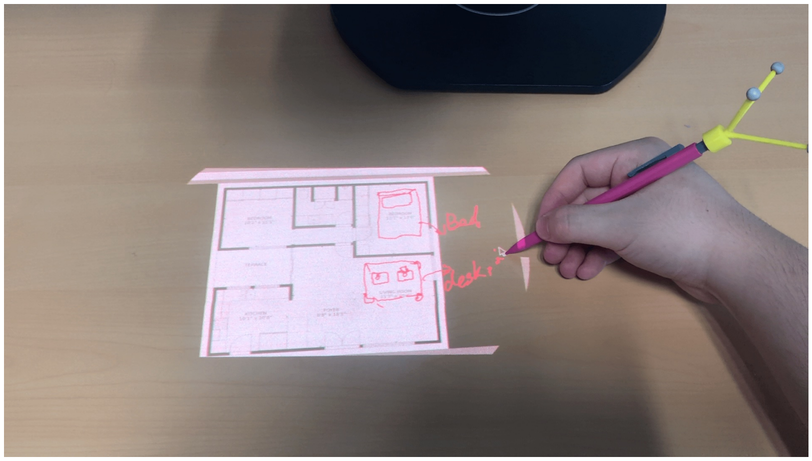
To relax, Alice plays music and moves the player to the surface using LuxAR. To choose a new song, she adjusts the lamp's height to explore different interfaces: at a medium height, she sees a list of songs with covers (Figure 3.15a), or she can use buttons to loop through the list when the lamp is low. To share music publicly, she moves the player window to a nearby sound speaker, which offers music controls for volume and song selection (Figure 3.15b). To enhance the environment, Alice can also place the music player on the ceiling to create ambient visualizations that sync with the music (Figure 3.15c).

(b) Drop the music player to a speaker by the lamp can augment the speaker

(c) Visualizing the music to enhance the environment when pointed at the ceiling

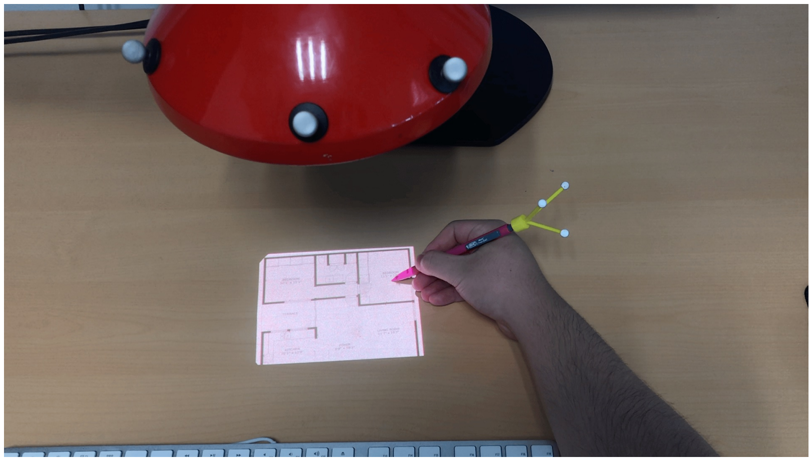
Figure 3.15: Pointing the lamp at a music player to augment the physical environment.

Drawing Annotation Tool

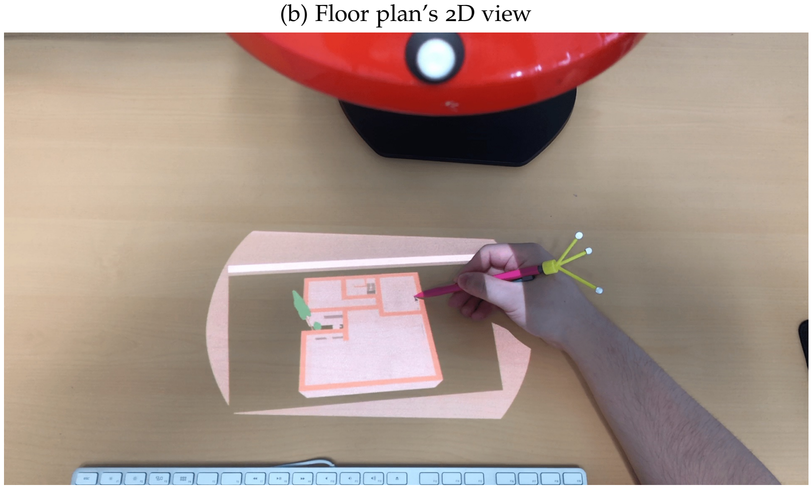


(a) Annotate the floor plan with a pen under the lamp

Later, LuxAR assists Alice with annotation tasks, such as drawing on a floor plan. She can move her digital workspace to a surface by pulling a floor-plan window from the monitor. Alice can switch between different floor plan views by changing the lamp's height: a medium height reveals the entire floor (Figure 3.16b), lowering it unveils a bedroom-scale view, and raising it presents a 3D-rendered floor (Figure 3.16c). Alice enhances her workflow by annotating directly on the surface with a pen, combining the benefits of digital and physical interaction (Figure 3.16a).



(b) Floor plan's 2D view



(c) Floor plan's 3D view

Figure 3.16: Annotate the floor plan with a pen by lamp height adjustment and when the lamp stays stationary.

3.5 USER STUDY

We conducted a two-phase user study to evaluate our system with the proposed interactions. In Phase 1, we evaluated participants' understanding and execution of these interactions. In Phase 2, we invited them to explore various scenarios. We further collected their feedback to identify ways to improve our system and enhance the current desktop experience.

3.5.1 *Participants*

We recruited 12 participants from a local institution, aged 20 to 33 (3 identify as women). 11 participants were right-handed and 1 was ambidextrous. 5 participants answered "yes" and 2 "maybe" on whether they have experience with architect lamps, and 2 reported using such lamps daily and weekly respectively. The study took approximately 60 minutes.

3.5.2 *Procedure*

The study consisted of three stages and employed a think-aloud protocol encouraging participants to provide immediate commentary. All comments and feedback were audio-recorded for subsequent analysis. The emphasis was on participant feedback and experience with LuxAR to extend desktop environments, rather than assessing task completion speed or efficiency.

3.5.2.1 *Introduction*

After reading the information letter and signing the consent form, participants were invited to adjust the seat and have a comfortable setting in the desktop space (Figure 3.11). Then, they were introduced to the study agenda, how our system was built, and how it could be used to manipulate virtual windows on the physical display or surfaces. Participants were allowed to manipulate the lamp and other devices to get a sense of the system.

3.5.2.2 *Phase 1*

In the first phase, participants engaged in application windows that display generic information and manipulated them within the desk space using the lamp. Each window featured a central icon, two navigation buttons to change icons, and a slider to adjust the background color. Additionally, windows provided information about the names of the object and the surface the lamp pointed at, and a distance indication categorized into three levels (small, medium, and high). Participants were assigned 15 tasks, which included activities such as changing icons and background colors on the display and surfaces using the mouse or touch and moving windows across surfaces and displays using the lamp. Following the completion of the task, the participants assessed physical and mental demands on a 10-point scale (1—very low, 10—very high). These two scales are a subset of the NASA TLX, using only two focused our investigation and reduced the burden for participants. Subsequently, open-ended interviews were conducted to collect feedback in depth on the manipulations and interactions. See also the accompanying appendix.

3.5.2.3 *Phase 2*

In the second phase, participants used the interaction techniques acquired in the previous phase to explore the applications and scenarios described in Section 3.4. Following the exploration, participants were interviewed to share their experiences with the applications, offering insights into potential system enhancements or scenario improvements.

3.5.3 *Data Collection*

We observed and noted how participants used our system to complete tasks during the study. We transcribed audio recordings and extracted comments from each question, as well as comments during each interaction. These notes were organized by question in a visual diagramming tool. To analyze the data, we applied a content analysis approach [73] to examine participant feedback on the direct manipulation of the lamp and associated interaction designs. Specifically, the main author analyzed the semantics of the data, grouped them by question, and then assigned them to each interaction design category. This analysis could help identify potential design considerations for designing explicit and direct manipulation interfaces for lamp-based systems that extend desktop computing.

3.5.4 *Results*

Participants (9 out of 12) found the system engaging and highlighted its ability to extend the current desktop space as *"...you are not trapped in your screen and have more spaces"* [P8] and its *"concepts are novel and practical"* [P10]. Participants also highlighted areas for improvement, particularly in the mechanical design of the lamp. They pinpointed specific challenges related to certain axes and recommended that enhancing degrees of freedom, coupled with increased lubrication in some joints, could further improve their experience. These observations were reflected in the self-evaluated mental demand ($M = 3.33$, $SD = 2.01$) and physical demand ($M = 5.42$, $SD = 2.27$).

3.5.4.1 *Intuitive manipulation across surfaces, but angles matter.*

The reposition of windows on the table was considered straightforward, but challenges arose when moving them to the wall or ceiling. Participants noted that such manipulation *"requires a lot of body movement"* [P4] and that *"[manipulation] angles matter"* [P7] with awkward wrist positions for wall and ceiling interactions. As commented by P12, *"For some angles, it may be relatively easy to handle, but for others, it may not be so straightforward ... I feel that it may be related* *to how I grasp the lamp."* We also observed that in Phase 1, six participants accidentally dropped a window behind the monitor during table-to-ceiling manipulations, an issue not observed in other across-surface manipulations, and none occurred in Phase 2. This showed how direct manipulation angles potentially impacted content repositioning to extend desktop computing onto diverse physical surfaces.

Participants were also positive about using the lamp to transfer virtual information between the monitor, physical surfaces, and mobile devices, which was described as *"neat from a digital point of view"* [P8], and *"context-wise interesting"* [P7]. However, concerns arose regarding the repositioning from a different source to a phone. Some participants saw it as a means to *"transfer data from computers to phones"* [P2], while others found the window's shape inconsistent on the phone, prompting questions about its purpose (P5, P7). P10 believed it existed a discontinuity between surfaces and devices when the same information was split across them, as shown in Figure 3.14c and *"...the information should be contained in one space."*

Only four participants noticed the visual change of the style on the lamp display as *"area is different"* [P7], but *"I did not know why it is happening"* [P4, P12] and three participants used the radar mode to find missing windows during the interaction without noticing other modes (hover, focus, and drag). Participants' focus was mainly on the manipulated windows, as *"I don't notice it and focus on the app as long as I can see"* [P6] and *"... I think I perform the manipulation earlier than noticing the visual effect"* [P11].

3.5.4.2 *Surface adaptation influenced by content and context.*

The adaptability to surface changes was prominently influenced by the content and contexts of manipulated applications, particularly exemplified in scenarios such as the calendar and music player. In the calendar example, P7 and P10 noted privacy concerns when asked to share the calendar with the public on the wall, but none raised this for the music player example. Besides, the ceiling was deemed unsuitable for the calendar by participants who rarely looked at it, as *"I really don't look at the ceiling at all and won't notice it"* [P3]. However, the adaptation was well received for the music player by participants: P1 expressed enthusiasm about the music player's ceiling visualization, envisioning it as *"super cool"* for wall visualizations while working, and P3 thought the visualization created a mood in an office, *"like in concert or club setting"*. Moreover, the adaptation should extend beyond the content within the window to encompass window sizes, particularly on larger surfaces, as P3 conveyed dissatisfaction with uniform window sizes on both a table and a wall.

3.5.4.3 *Varied opinions on interactions based on height adjustment.*

Similarly to the results in §3.5.4.2, participants perceived the ability to change details and content through the manipulation of the lamp's height differently based on tasks and application contexts. Initially, eight participants were perplexed by this interaction in Phase 1, with P4 expressing *"this provided new possibility to interact with the lamp but did not provide clear use cases"*. However, in Phase 2, participants favored the floor plan example to illustrate this interaction, with positive comments such as *"Floor plan is cooler... it makes more sense to me"* [P8], and P2 adding, *"You can see different aspects, and this makes sense for the architect lamp."* For the calendar example, while participants understood the concept, the limited lamp display size prevented them from reading the information, showing only a restricted part of the window when the lamp reached a lower level. Comments included *"It is more private, but not convenient. I don't like the view to be so limited ... have to go back and forth to see everything"* [P3], *"It is disturbing to move the lamp to explore the calendar because only parts of the calendar are shown"* [P2], and *"The closer the lamp is, the less information is shown. I have to move"* [P7]. Moreover, adjusting the height posed a challenge when pushing the lamp to a lower level and maintaining that pose. P8 highlighted this issue, stating, *"The concept is easy to understand. Just the mechanical parts... you could not keep it low."* This observation raises ergonomic concerns when designing interactions based on surface proximity for lamps.

3.5.4.4 *Object augmentation enjoyable, but tricky to trigger.*

Participants positively responded to the augmentation of physical objects like the clock and speaker, finding it interesting and enjoyable, such as *"music player is funny with the speaker"* [P2]. While augmenting the phone with additional views was favored, as *"this is nice to have extended views for phones"* [P8], opinions about using the lamp to complete the information transmission from different surfaces to the phone varied, as mentioned in §3.5.4.1. While participants enjoyed the augmentations, they encountered challenges when triggering them and sought the help of the study facilitator. Although they grasped the concept of picking up a window and moving it to objects, they struggled with the drop action, specifically, where and when to release the button. For instance, some participants were observed to wait for the interaction to activate automatically or try to induce it by pressing the lamp downward. As the release action was non-intuitive, P4 and P8 suggested adding visual cues and feedback around the physical objects to guide the release action. P12 also suggested increasing the boundary for object detection.

3.5.4.5 *Challenges with detached interactions*

Participants were tasked with changing the icon and background color of the window using the mouse, touch and mobile device. Although the use of the mouse on various surfaces was generally perceived as straightforward, several participants noted challenges in tracking the cursor as the lamp moved across surfaces (P2, P7, P11) or in low resolution within the lamp display (12) and P8 reported misaligned mouse coordinates between the lamp display and the application due to a tilted lamp angle. Only half participants were positive about touches on surfaces. P4 preferred to *"perform pinch with touches* *[on surfaces]"* but for *"simple actions, I would go back to the mouse and keyboard"*. P5, who was used to touchscreen interfaces, enjoyed interacting with the same content on physical surfaces. In contrast, P9 noticed a latency for touches on surfaces and P11 felt that *"...the sensitivity is not the same [for touches on screens and surfaces]"*. The introduction of the pen tool during the floor plan example did not elicit additional feedback from participants. They focused on how the height adjustment of the lamp influenced the content, implicitly reflecting the naturalness of using the pen with LuxAR.

3.5.4.6 *Other Possible scenarios*

Participants suggested various applications for LuxAR in both phases. In Phase 1, eight participants envisioned a collaborative environment, including multiplayer games and information sharing across devices. Two focused on information management with applications for maps, calendars, and drawings. One participant suggested using virtual objects for decoration, revealing them with the lamp and one believed that our system could be quite useful for parts assembly scenario when instructions can be superimposed onto objects and users can use both hands to achieve assembly tasks. In Phase 2, two participants proposed context-aware applications. P1 recommended a smart home control concept, suggesting, *"...put everything on the clock to show time-related events..."*. P4 suggested to *"...show the content depending on where the lamp is actually located... closer to the bed... closer to the desk or in the kitchen"*. P9 believed that it can be used for a football match such as *"pointing at a player and I can see the player's statistics for this match"* and P12 planned to *"show API documents when I point at the code in a window and then move the documents on other space"*. Two participants suggested replacing application windows with information windows to aggregate data from various sources.

3.6 Discussion

Throughout our two-phase study, participants demonstrated a clear understanding of how to manipulate LuxAR to explore different scenarios and grasp the interaction designs centered on an architect lamp. They found it intuitive to use the lamp as both an input and output device to extend and augment desktop computing environments. This ease of use may be attributed to the design of the direct input, where our lamp is employed to move content on surfaces, akin to mouse input in desktop computing: button-based interactions closely resemble mouse clicks, while physical lamp manipulations mirror mouse movements on a surface. Although some participants faced challenges in operating the button in specific lamp orientations (*e.g.* moving the window from the table to the ceiling), mental demand was low overall, and none of the participants had difficulty learning the manipulations and interactions.

Our results also showed that participants had varied feedback on how projected content adapted to lamp heights. For instance, participants favored this interaction to draw on the floor plan compared to using it with a calendar.

We also noticed that the ceiling was less favored compared to other surfaces. These observations suggest that the adaptation of virtual content to dynamic surfaces and the direct manipulation of a SAR system could be surface- and context-dependent or a combination of both. The limited use of the ceiling may also stem from the lack of a tool for displaying information on it. These open up a broader design space for directly manipulated SAR, offering potential avenues for future research.

Although understanding the interaction is straightforward, uncovering all available interactions can be a bit tricky. Our results revealed that while manipulating the system, participants primarily focused on the manipulated window, so the majority of participants neglected the visual changes in the lamp display, which help to signify the interaction state of the window. Meanwhile, they faced challenges triggering augmentations on physical objects, highlighting the necessity to amplify visual feedback around the physical objects when they were under the lamp. This indicated that rather than changing the projection display style, placing obvious interaction indicators around virtual windows or physical objects was more visually perceivable.

3.6.1 *Design Implications*

Our findings indicate that using an architect lamp as both input and output device is intuitive, but some aspects of its manipulation may be challenging. Additionally, designing attached and detached interactions for direct manipulation systems requires careful thought to accommodate different contexts. We revisit our design goals and suggest further considerations for explicit and direct manipulation of lamp-based systems, and SAR more broadly, when integrating desktop computing into physical spaces.

Mechanical Designs

Compared with previous efforts, such as the Lantern [135] and LuminAR [139], our system explores an architect lamp's direct manipulation potential to interact with the content across different spaces **[DG1, DG2]**. It also leverages the unique mechanical structure of the lamp to offer stable positioning and persistent augmentation without requiring continuous user engagement. This allows detached interactions and inputs from the lamp's manipulation, such as mouse, touch, pen and devices **[DG4]**. However, as the lamp is used as a direct manipulation input, participants encountered challenges in operating the lamp, desiring more degrees of freedom (DoFs) to address issues with specific axes. It is also noted that architect lamps are predominantly designed to face downward. Consequently, unlike handheld systems [22, 45, 104, 249, 260] employing the torch form, certain axes and placements of our system were challenging to execute during the study due to the lamp not being explicitly designed for such movements. Additionally, the fixed placement of the lamp on the desk limits mobility.

To address these concerns in architect lamp-based SAR systems, potential solutions include attaching wheels for easy desk movement and replacing current joints with spherical joints or 6-Axis force-torque sensors with locking capabilities for improved flexibility in positioning and orientation **[DG1, DG2]**. Future designs should balance the trade-offs between mobility and stability **[DG1, DG4]**, considering the specific form factors of SAR systems.

Attached Interactions with Spatial Awareness

Direct manipulation of the lamp requires some degree of spatial awareness to effectively interact with content. When moving a window across surfaces, we show how the content can adapt, offering a new style of interaction compared to other direct manipulation pro-cam systems [32, 45, 249]. Explicit control of augmentation also allows users to decide where to display information and what information to display. This makes interaction more focused and avoids being distracted by augmented information displayed elsewhere with a multiple pro-cam system simultaneously displaying content on many nearby surfaces. However, this may also limit people from browsing information outside the chosen augmented area (*e.g.* consider P3 when reading the calendar) and could slow down workflows spanning multiple sources of spatial information. Note many single pro-cam systems face the same challenges [32, 45, 116, 249].

Our results suggest that ceiling usage was less favored, possibly due to users finding it irrelevant when seated. Despite this, the ceiling could serve as ambient cues for conveying messages **[DG2, DG3]**, *e.g.* a countdown event could inform Alice's colleague. Moreover, while Tomitsch et al. [226] suggested ceiling use for information visualization and Satkowski et al. [203] recommended low visual complexity content on the ceiling in an HMD setting, most prior manipulation-based SAR systems did not consider the ceiling as an interaction space. This creates an opportunity to explore ceiling interactions, with LuxAR serving as a tool for such exploration. For example, parents and children can lie on the bed to use architect lamps around their table to control multiple characters, as in [248, 249] or tell bed stories, as in [247].

Our system employs three proximity levels for information granularity [45, 222]. Content remains consistently sized within each level, zooming with lamp height adjustments and crossing proximity levels alters the context displayed. Our findings show that the participants preferred to use this interaction to draw floor plans, while they encountered limitations in searching and reading information on the calendar. This suggests a potential context-dependent adaptation for height-adjusted interactions in the design of direct manipulation-based SAR systems — less suitable for content with high text density **[DG1, DG2]**. Additionally, our prototype employed consistent values for desks and walls, omitting this interaction for the ceiling due to the available manipulation axes. Future research could explore this area, broadening this interaction based on varied combinations of manipulations and surfaces with a more flexible manipulation-based lamp SAR system.

It is also important to effectively inform users about the interaction state and guide them on when, where, and how to initiate interactions, especially when interactions are attached to the direct manipulation of the system itself

[DG1]. While we propose changing the lamp display style to indicate the interaction state of a content, future work should consider applying the visual signifiers to the virtual content and physical objects and devices directly and explore how different designs and visual significance could better guide users in direct manipulation-based interactions.

Detached Interactions with Stationary Lamp

When LuxAR left untouched, it maintains the augmentation and allow users to interact with the content with other inputs **[DG4]**. Although user can directly annotate information by touch and pen on the table, they are not suitable for the walls and ceiling, and the mouse is generally preferred, but participants encountered tracking challenges. While cursor loss is a common issue [142], enhancing cursor visibility during manipulation, as previously discussed, is a potential solution for SAR systems using direct manipulation. Additionally, considering coordinate remapping between the manipulated display and the content can assist users in performing effective mouse interactions across different surfaces. While a mobile device support touch inputs, but its high resolution touchscreen make it a better holder for content. Therefore, future design on Temporary View could split up the information and control: placing information on a device and moving control and interactions onto surfaces lighted by the lamp.

3.7 CONCLUSION

In this chapter, we present LuxAR, a direct manipulation SAR prototype that uses an architect desk lamp as an input and output tool, extending desktop computing to various surfaces in the desk space. By leveraging the lamp's manipulation capabilities, users can manage virtual content across physical displays, objects, surfaces, and mobile devices, interacting with it through the lamp and four other inputs (RQ2). More importantly, our augmented lamp establish bidirectional connections across *Environment* dimensions through its physical manipulation (RQ1) — virtual content can stay within the screen, be placed on a surface or an object when the lamp is not manipulated, moved to other surfaces when the lamp is manipulated, and back to the original *Environment* dimension given users' activities. Meanwhile, the content adapts the physical environment accordingly during the lamp's manipulation. Our user study demonstrates that participants understand the design of our lamp system and use it to explore three demonstration examples, and showcases its potential to enhance the capabilities and accessibility of desktop computing environments (RQ3).

4 INVESTIGATING COUPLED AND DECOUPLED TARGET ACQUISITION OF DYNAMIC PEEPHOLE ON AN AUGMENTED DESK

When users manipulate the augmented lamp to re-position the virtual display, it creates a dynamic peephole interaction space for the augmented workspace. As one hand is dominated by lamp manipulation, the impacts of control mechanism on target acquisition for the dynamic peephole interactions on the augmented surfaces are unclear. In this chapter, we conduct a controlled experiment in Virtual Reality to understand the impact of two control mechanisms for target acquisition tasks in a dynamic peephole display.

4.1 INTRODUCTION

Peephole interaction refers to a two-step process in which an object of interest hidden within a virtual space is first brought into a viewport (commonly referred to as a *peephole*) before an interaction is performed like selecting a target [83, 84]. A peephole can be physical [263] or virtual [120], but the first step of interaction always requires a search phase in order to bring it to the view. For instance, annotating a paragraph in a long document on a desktop computer is a type of peephole interaction since it requires scrolling to locate the paragraph of interest. Note that the first step to scroll is not necessary if the text is already visible, making it a non-peephole interaction.

Based on how the object of interest is brought into the peephole [155], the interaction is *static* or *dynamic*. A static peephole disregards the physical movement of the peephole and the virtual space is manipulated "behind" the peephole to bring information into view. For example, scrolling a document in the example above. In contrast, a dynamic peephole uses the physical movement of the peephole itself [45, 120] to search the virtual space. This is exemplified by a peephole created by a handheld-projector [45, 56], the position of a handheld display [122, 184, 228], or the field of view of a head-mounted display (HMD) [57, 216].

This raises the question: should peephole manipulation and task execution be integrated into a single device (*coupled*), such as revealing digital information on physical surfaces and selecting it by pointing [56] or crossing [45], or should each task be handled separately using different inputs (*decoupled*), such as touch devices [57], gaze or controller [216]?

Both coupled [85, 120, 121, 248, 249] and decoupled [45, 218] control conditions have been leveraged to design dynamic peephole interactions. Identifying their differences can help improving the design of interaction techniques. Previous studies have investigated 1D reciprocal pointing tasks [47, 216], where the direction of a target is known, and the search phase is limited to moving towards the target. However, dynamic peephole interactions are typically 2D [45, 218, 249] and can extend across multiple surfaces [42, 56]. This introduces very different physical characteristics when manipulating the peephole. Searching for the target is also more difficult since both direction and distance are unknown. Initial investigation into 2D dynamic peephole pointing have included visual guidance [57], or presented mixed analysis on targets both within or outside the peephole [133, 215].

In this work, we examine *whether target acquisition in a dynamic peephole on augmented surfaces should be coupled or decoupled*. We setup our study in a desk workspace and created a virtual lamp [56, 139] as a peephole on the desk. Participants were asked to move the virtual lamp to reveal targets on the desk and then select targets using both coupled (the peephole centre for selection) and decoupled techniques (direct touch for selection). We found subtle differences in accuracy, total time and search time between the coupled and decoupled methods; however, the coupled method was significantly faster in acquisition time. Notably, during the search phase with the coupled technique, participants moved the peephole faster but revealed targets later, while the decoupled technique led to slower peephole movement but an earlier target reveal. Participants favoured the coupled technique for its convenience and reduced physical demand, but preferred the decoupled technique for accuracy. This was supported by how participants manipulated the peephole: the coupled involved minimal body movements with mainly wrist rotation, while the decoupled encouraged more dynamic, two-handed exploration and coordination. Both Cao, Li, and Balakrishnan's CLB model [47] and Huber, Steimle, and Mühlhäuser 's HSM model [107] proved inadequate for this 2D task on a single surface, but a modified HSM model incorporating target width yielded improved results. These results suggest design guidelines for controlling dynamic peepholes in an augmented desk workspace. In summary, we contribute new findings for the impact of coupled and decoupled target acquisition on dynamic 2D peephole pointing in an augmented desktop space.

4.2 RELATED WORK

In Chapter 2, we explain that the direct manipulation of an SAR system creates a dynamic peephole display. Here, we present a board content of peephole interactions.

While peephole interaction and magic lens interaction [29, 30] are often discussed together, the latter provides visual cues outside the peephole to guide users to targets. Our work focuses on peephole interaction, specifically dynamic peephole interaction.

We begin by reviewing previous research that leverages the dynamic peephole for interaction to highlight its importance and growing use in diverse environments. This is followed by studies on target acquisition within the dynamic peephole, highlighting the impact of various factors on this interaction and the lack of understanding of control mechanisms in dynamic peephole interactions. We summarize relevant previous work in Table 4.1.

4.2.1 *Interaction with a Dynamic Peephole*

Although static peephole is more common in different interactions, such as map navigation in mobile devices, dynamic peephole interaction has shown its advantages in different tasks. For instance, Mehra, Werkhoven, and Worring showed that a static peephole takes more time to discriminate lines in a desktop interface than a dynamic peephole [155]. Rohs et al. found that dynamic peephole techniques have advantages in search time and exploration space for mobile map navigation tasks over static peephole using joystick interactions [197]. Similarly, while the dynamic peephole is slower than the static peephole for map navigation on a smartwatch [122], it is preferred and facilitates larger navigation areas. Hürst and Bilyalov demonstrated benefits of dynamic peephole for navigating 3D virtual environments with mobile devices [108]. Additionally, dynamic peephole not only improves operator performance in recall information [183], but also improves the performance of observers when multiple users collaborate in a navigation task [121].

These beneficial characteristics have motivated researchers to design novel techniques and systems based on dynamic peephole interactions. For example, Yee developed a spatially aware mobile device to annotate virtual information in a large workspace with two hands [263], and Spindler, Stellmach, and Dachselt designed a system to allow users to move a paper over a projected skeleton to examine different layers of information [222]. People could leverage this interaction to spatially transfer content across displays and surfaces [35, 144, 171], perform multi-level selection on spatial items [52, 221], and examine spatial visualization [41, 228] by moving and interacting with tracked mobile devices. Practitioners used mobile devices as the viewport to the physical world, and developed applications like Pokémon Go, to create mixed-reality interactions. More recently, Chen et al. designed an augmented lamp to navigate and manipulate information across different surfaces around the desktop environment [56]. As dynamic peephole interactions become more accessible, it is important to investigate factors that impact the performance of the dynamic peephole for the design of interaction techniques.

4.2.2 *Dynamic Peephole Target Acquisition*

Previous studies have investigated different factors that impact dynamic peephole target acquisition in various environments. Among these factors, peephole size is one of the most frequently studied. Cao, Li, and Balakrishnan examined various sizes of peephole in a 1D pointing task, in a desktop environment, proposing a modified Fitts' law to model the interaction [47]. Follow-up studies [76, 120, 216] validated this model in different contexts, but still for 1D pointing tasks. Rädle et al. noted that while peephole size affects performance, bigger is not always better [185]. They found that a tablet-sized peephole was a sweet spot for performance and preferences with a large external wall. Müller et al. found that with the same peephole size,

	Interaction Space	Type	Dimension	Coupled	Decoupled	Size	Prior Knowledge	Visual Cues	Orientation
Our Work	**Augmented Surfaces**	**Acquistion**	**2D**	✓	✓				
Cao et al. [47]	Desktop	**Acquistion**	1D	✓	✓	✓	✓		
Sidenmark et al. [216]	Immersive 3D Space	**Acquistion**	1D	✓	✓	✓	✓		
Chen et al. [57]	**Augmented Surfaces**	**Acquistion**	**2D**	✓	✓			✓	
Huber et al. [107]	Physical Paper strip	**Acquistion**	1D	✓					
Kaufmann and Ahlström [120]	**Augmented Surfaces**	**Acquistion**	1D	✓		✓	✓		
Forlines et al. [85]	**Augmented Surfaces**	**Acquistion**	1D	✓					
Ens et al. [76]	Cave Display	**Acquistion**	1D		✓	✓	✓		
Araki and Komuro [9]	**Augmented Surfaces**	Search	**2D**	✓					
Rohs et al. [43]	Mobile Display	Search	**2D**	✓				✓	
Kaufmann and Ahlström [121]	**Augmented Surfaces**	Search	**2D**	✓					
Rädle et al. [185]	Wall Display	Search	**2D**	✓		✓			
Müller et al. [162]	Mobile Display	Search	**2D**		✓				✓
Mehra et al. [155]	Desktop	Line-length discrimination	1D	✓					
		Task		**Control Conditions**		**Other Peephole Factors**			

Table 4.1: Previous dynamic peephole interaction compared to our work (sorted by relevance, see text for comparison criteria).

the peephole orientation does not significantly affect navigation performance or spatial memory.

Prior knowledge is often referenced as the spatial location of an object of interest in acquisition [47, 76, 216] and navigation tasks [106, 183]. More specifically, it comprises two aspects: the direction and the distance to a target. Cao, Li, and Balakrishnan studied this factor and found that people with prior knowledge of target locations show faster, and more accurate performance compared to those without prior knowledge [47]. Other studies confirmed this result in augmented surfaces [120], cave displays [76], and immersive 3D space [216]. However, since these tasks were one-directional, and targets had the same height as the workspace, participants were aware of the direction and simply had to move left or right to acquire targets, similar to an aimed movement. This may explain why the standard Fitts' law [82] was found to be valid [45, 120]. To address this concern, Huber, Steimle, and Mühlhäuser designed a physical apparatus that avoids offering directional hints and proposed a model based on physical body movements [107]. However, their model does not consider the width of the targets, and may not be generalized to a 2D context where both the direction and distance to a target are unknown and users have different physical characteristics for operating the peephole.

Compared to other factors, the (de)coupling of peephole control and target acquisition has received less attention. Cao, Li, and Balakrishnan found that in a desktop setting, a coupled mechanism led to faster interaction than the decoupled control [47] while Sidenmark et al. reported that with the HMD viewport as the peephole, decoupled methods (gaze and controller) showed advantages over the coupled method in terms of accuracy and speed in VR [216], but both their tasks on the workspace were 1D, and provided directional cues. In addition, these tasks had limited spatial complexity and used a simplified motor control. On the other hand, while Chen, Katsuragawa, and Lank examined this interaction across multiple surfaces, the visual guidance eased the search phase, as it did not require participants to memorize the spatial locations of the targets (*i.e.* prior knowledge), and transformed the task into an aiming interaction [57].

It remains unclear whether target acquisition in dynamic peephole as a 2D task should be coupled or decoupled. We want to examine this question without visual guidance [57] and without requiring participants to find targets by following one direction. Instead we want participants to be aware of its previous relative location during the interaction, to better understand how interaction techniques impact both search and acquisition behaviors.

4.3 METHODOLOGY

In order to investigate the performance of (de)coupled dynamic peephole target acquisition, we conducted our studies in a reconstructed virtual desk workspace using a head-mounted display (HMD). A typical desk workspace includes various surfaces, such as a table, monitor, and wall, making it an ideal setting for revealing information through a peephole and adapting to coupled or decoupled techniques depending on the distance [42, 56, 184]. VR can be used to precisely control experimental conditions, eliminating physical distortions of the peephole's shape and size that often occur in projector-based systems due to surface irregularities and environmental lighting [9, 45, 56, 120]. This control minimizes confounding variables, ensuring more reliable data. Besides, as dynamic peephole interaction typically involves frequent physical body movements, we wanted participants to remain seated to improve comfort and reduce potential fatigue and motion sickness, particularly in VR.

4.3.1 *Creation and Mapping of Virtual Surfaces*

When simulating the problem in a virtual environment, we create virtual surfaces that mimic their physical counterparts, enabling users to interact with targets as if they were touching physical surfaces. To achieve this, we place reflective markers on various surfaces, such as tables and walls, and utilize optical tracking systems like OptiTrack and Vicon to register these markers as rigid body objects. Their locations and orientations are then saved locally. However, the coordinate system used by the optical tracking systems differs from that of head-mounted displays (HMDs), particularly when using devices such as the Oculus Quest 2, which resets its coordinates at startup. To resolve this issue, we use a transformation matrix to map the position and orientation of physical objects to virtual objects. To obtain this transformation matrix, we anchor a reflective spherical marker on a physical controller and create a virtual sphere at the corresponding location in the virtual environment. A calibration procedure is performed when the system starts, as detailed in Algorithm 1. The resulting transformation matrix is then applied to update the position and orientation of virtual objects based on the physical environments. This approach also allows us to work with any optical tracking system and virtual environment created by HMDs that utilize physical controllers.

Algorithm 1 Compute Transformation Matrix

Initialize empty lists *Src*, *Dst*

//Collecting matching points for calibration
while $|src| \leq 500$ **do**
  Add positions of physical & virtual spheres to *Src* and *Dst*
end while

//Normalize point clouds to avoid scaling factors
$Src_{norm}, Src_{transM} \leftarrow \text{Normalize}(Src)$
$Dst_{norm}, Dst_{transM} \leftarrow \text{Normalize}(Dst)$

//Perform calibration using generalized iterative closest point technique
$Src_{norm}, Dst_{norm} \leftarrow \text{EstimateNormals}(Src_{norm}, Dst_{norm})$
$Ret_{norm} \leftarrow \text{GeneralizedICPRegistration}(Src_{norm}, Dst_{norm})$

//Applying the scaling factors to construct the transformation matrix
$Ret \leftarrow Dst_{transM}^{-1} \cdot (Ret_{norm} \cdot Src_{transM})$
return *Ret*

4.3.2 *Manipulate Peephole and Acquire Target*

We use the flashlight (lamp) metaphor [56, 186] to create a peephole display on virtual surfaces. A virtual lamp is placed in the environment, and the peephole display is cast on the virtual surfaces based on the lamp's pointing direction. The virtual lamp can be manipulated with 6 degrees of freedom (DoF). To minimize the impact of hand jitter on the peephole display when grasping the lamp, we filter the lamp's direction with a 1€ filter [50]. The peephole display itself is represented by a rectangle with a solid black outline, displaying only the intersection area between the target and the display.

To acquire a target, users start by grasping the virtual lamp with their non-dominant hand. For that, they place the controller within the virtual lamp, and press and hold the trigger button. Then, they manipulate the lamp to move the peephole display on surfaces to search for a target. Users release the lamp by releasing the trigger button to stop moving the peephole. When a target is partially revealed within the peephole, it can be selected, and participants can use the corresponding technique (coupled or decoupled, depending on condition) to select it.

For the coupled technique (Figure 4.1), we place a cross-hair at the center of the peephole display, used as the cursor for selection. When the cross-hair intersects with a target, the target is highlighted, and users can press the index trigger of the controller they are grasping to acquire it.

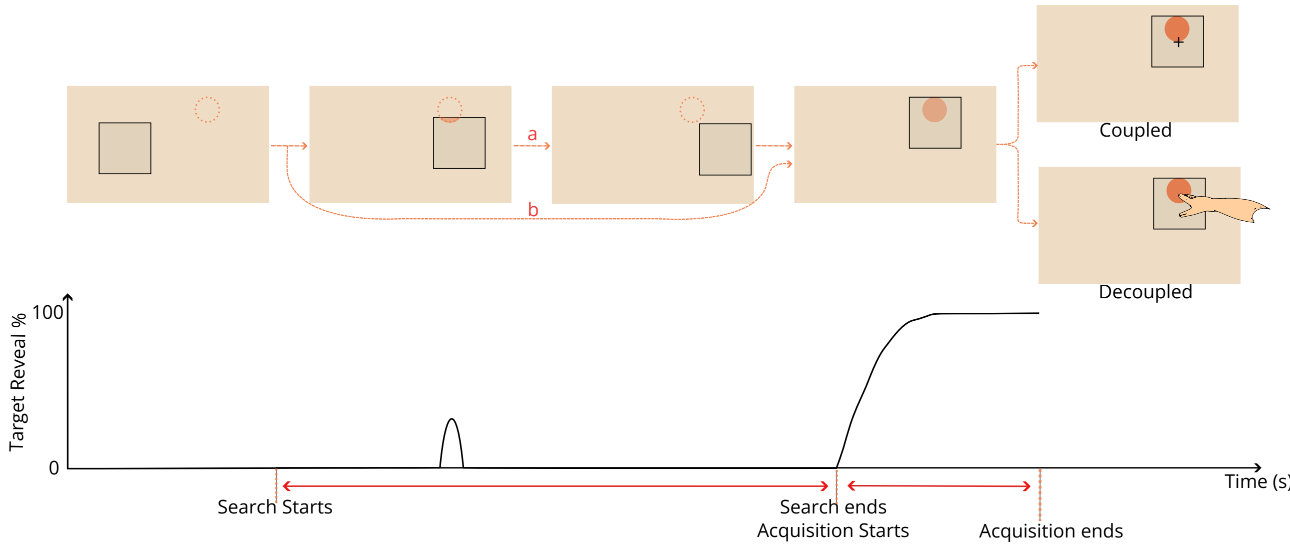


Figure 4.1: The timeline of searching and acquiring targets for the dynamic peephole target acquisition with the coupled and decoupled methods. During the search for a target, users might (a) accidentally reveal it but not be aware of it, or (b) directly find it.

For the decoupled technique (Figure 4.1), the control of the peephole remains the same as in the coupled technique, but without the cross-hair. Instead, users perform direct touch with their dominant hand's index finger to select a visible target. The target is highlighted when it is revealed and the finger hovers over it, and users tap on it to acquire it.

4.3.3 *Decoupled Control with Direct Touch*

While hand tracking is often supported in HMDs like Quest 2, it can be unstable and prone to producing false actions [143] and high position deviation [207] with direct touch. The most recent work by Bérard [43] used an optical tracking system to enable more accurate direct touch on physical surfaces. It proposed an optimization approach to model the fingertip as a sphere and minimize the tip-surface distance. Although this approach achieves submillimeter accuracy in touch offset, it reports relatively high false positives for touch detection. To achieve higher reliability, we propose a dual-module approach combining the ideas of Bérard's work (as a touch tracking module) and Masson et al.'s WhichFinger method [150] (as a touch detection module). For the tracking module, we create a ring mount with three sticks to attach three reflective markers and register them in the optical tracking system, which is subsequently used to obtain the fingertip position. The detection module operates independently of the tracking module, detecting a touch vibration signal on physical surfaces. We then combine inputs from both modules for reliable touch detection.

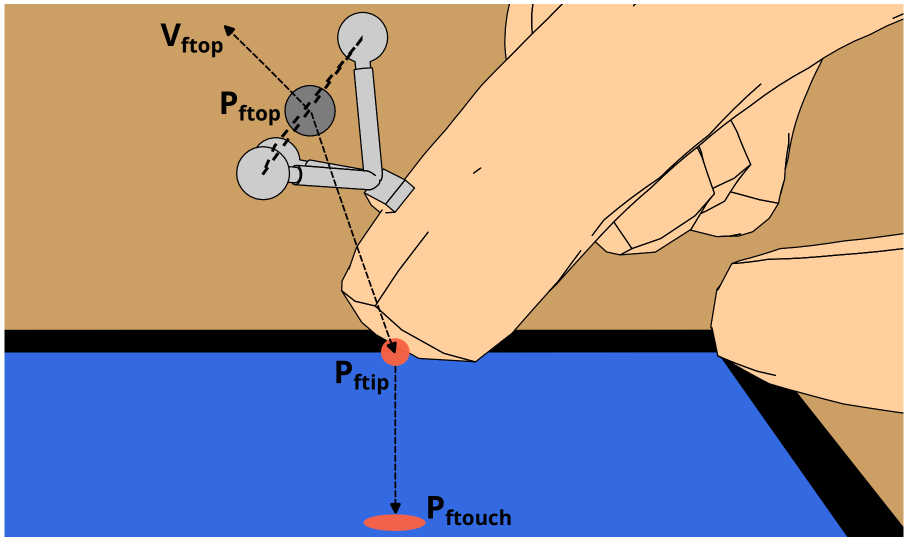


Figure 4.2: Optimization of finger tracking.

Touch Tracking

When the ring is worn in the distal interphalangeal joint of the index finger, the center of the ring is defined as f_{top}, the index fingertip is defined as f_{tip} and the orthogonal projection of fingertip on the surface is defined as f_{touch}. Instead of modeling the fingertip as a sphere and optimizing the distance between the tip and the surface for touch tracking and detection [43], we make the assumption that f_{top} will not move after the ring is worn so that the vector defined by f_{top} and f_{tip} is invariant at the local coordinate defined by the ring regardless of the movement of the index finger in the global coordinate defined by the optical tracking system. Therefore, we describe the following objective optimization function to estimate the position of f_{tip} based on the position and orientation of f_{top}:

$$\underset{R,s}{\text{minimize}} \quad ||T_{f_{top}}(P_{f_{tip}} - P_{f_{top}}) - s * (R \cdot T_{f_{top}}(V_{f_{top}}))||$$

The transformation matrix R and scalar s serve as optimization parameters. The function $T_{f_{top}}$ transforms the vectors from the global coordinate to the local coordinate. $P_{f_{tip}}$ and $P_{f_{top}}$ are positions of f_{tip} and f_{top} respectively, and $V_{f_{top}}$ denotes the up vector of f_{top} in the global coordinate, as illustrated in Figure 4.2.

To optimize R and s, we place a touchscreen (only for optimization purposes) on the table and use a tracked pen to mark the four corners of the screen. When the index finger touches the screen, we use the positions of the marked corners and pixel positions from the touchscreen to interpolate $P_{f_{tip}}$. We collect 500 points for $P_{f_{tip}}$, $P_{f_{top}}$, and $V_{f_{top}}$ and use them to optimize R and s. With optimized R and s in hand, we can now estimate $P_{f_{tip}}$ using $P_{f_{top}}$ and $V_{f_{top}}$ as reference points. In the virtual environment, a sphere is positioned at the predicted $P_{f_{tip}}$ for visualization purpose. An invisible circular area f_{touch} is then generated by orthogonal projection from f_{tip} to the nearest surface

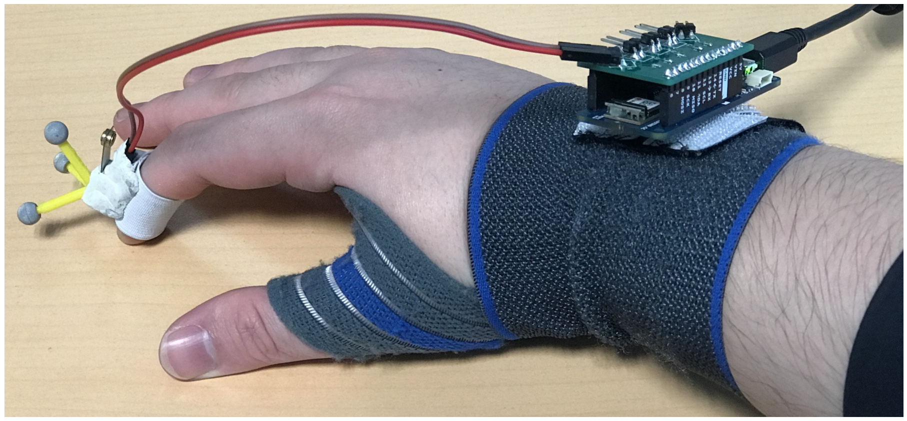


Figure 4.3: Prototype of dual-module touch detection method.

and used for interaction. The widths of the spheres at f_{tip} and f_{touch} are set to 1.8 cm [68, 89, 238] to accommodate different finger sizes across participants. For consistency, the size of the cross-hair in the coupled method is also set to this value.

Touch Detection

Touch detection is implemented independently of touch tracking, leveraging the WhichFinger method [150] for its high flexibility, low cost, and high accuracy. To detect vibrations, a vibration sensor is attached to the base of the ring, while an adhesive tape secures the Arduino Leonardo board using a wrist protection worn by the user. We again use the 1€ filter to smooth the vibration signals, apply an empirical threshold to filter high-voltage signals, which are subsequently merged if they fall within timeframe. As we only use one vibration sensor for touch detection and try to avoid false positives due to the ripple effect when detecting touch vibration, we discard subsequent touch signals if they appear within 300ms of a previous signalled touch event.

Touch Detection Accuracy

Our prototype is illustrated in Figure 4.3. We conducted a pilot study to evaluate the accuracy and stability of our approach by defining an area with four markers on the desk, which was then roughly divided into 9 subareas. In each sub-area, we placed a tablet and used a tracked pen to mark the four corners of the screen. Wearing our prototype, we performed random touches on the touchscreen with various touch poses. This process generated 105 samples of position values, interpolated using marked positions and pixel positions. Meanwhile, our prototype might generate a single detection event, several, or none each time a touch was performed. We used the samples from the touchscreen as ground truth and computed the true positives (TP), false positives (FP) and touch deviation (TouchDev) of our method. With an average TP of 99.8%, FP of 0%, and TouchDev of 3.7mm laterally and 2.5mm

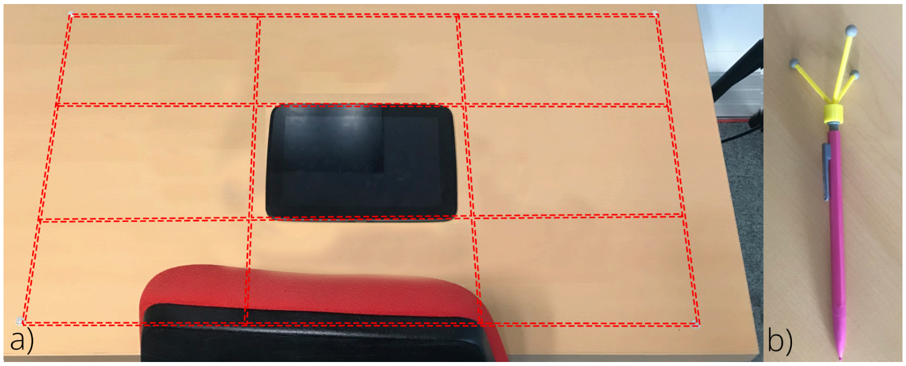


Figure 4.4: a) Pilot touch test setup on a desk (the surface is divided into 9 sub-areas) with b) the registration pen.

in depth within the tracking system's 3D coordinate frame, our prototype supports highly accurate touch detection.

	TP (%)			FP (%)			TouchDev (mm)		
Area	100	100	99	0	0	0	[3.4, 1.9]	[3.0, 2.4]	[6.5, 1.4]
	100	99	100	0	0	0	[2.0, 1.4]	[4.1, 4.7]	[6.2, 4.4]
	100	100	100	0	0	0	[3.8, 1.5]	[2.8, 2.2]	[1.1, 2.4]
Mean	99.8%			0			[3.7, 2.5]		

Table 4.2: Mean TP, FP and TouchDev in [x, z] axes of each sub-area in the tracking system's 3D coordinate frame.

4.3.4 *Data Collection*

Recall that our research question is whether target acquisition in dynamic peephole on augmented surfaces should be coupled or decoupled. We aim to address our question from two perspectives: the quantitative differences between coupled and decoupled techniques, and the impact of techniques on the preferences and kinematic profiles of the participants. Movement data for the peephole, the virtual lamp, and the fingertip sphere were logged as sequences of time-stamped, three-dimensional coordinates. Additionally, we logged all users inputs, system events such as controller button presses, finger-tapping actions, whether a target was revealed (including the corresponding reveal area percentage), and whether the selection was an error. The quantitative metrics of interest for targets were:

- Error rate — The number of selections that result in an error divided by the total number of selections.

- Total Time — The time between the previous correct selection and the next correct selection.
- Search Time — The time between the previous correct selection and when the target is found (see below).
- Acquisition Time — The time between the target being found and the next correct selection.
- Perceived Workload — Raw NASA TLX metrics [102].

We observed participants interaction strategies of dynamic peephole during the experiment, including postures of two hands and arms, body movement and wrist movements for searching and acquiring targets. We also collected their preferences and feedback of each condition at the end of the experiment.

Note that our work differs from previous studies on 1D tasks [47, 76, 120, 216] in the searching phase. In our context, the peephole might pass by the target several times without participants actually noticing it (see Figure 4.1 (a)), making it difficult to define when a target is considered found based solely on when it is revealed in the viewport for the first time (*i.e.* search time). To address this, we adopted a concept similar to Meyer et al.'s criteria for defining sub-movement of human motor performance [156]. We computed the target reveal percentage of a target using the amount of visible vertices. We then filtered the raw data on the percentage of target reveal using a zero phase filter and retrospectively analyzed the last ballistic movement between acquisition and target reveal to define the moment the target is found (Figure 4.5). We performed a reverse linear search to identify the first local minimum, whose value is below a threshold. We empirically set the threshold at 20% as it represents a good tradeoff between having a reveal percentage too small to have the target noticed by the user and too large so that the user is already in the acquisition phase. In practice, adjusting this value between 10% and 40% does not affect the results much, since we look for the local minimum close to the threshold value. This was confirmed by visual inspection of the results (see examples on Figure 4.5).

4.3.5 *Research Hypotheses*

Our research investigates *whether target acquisition in a dynamic peephole on a single surface should be coupled or decoupled*. We simulate a scenario where users engage in a reciprocal target acquisition task on a desk using a dynamic peephole. This scenario mimics the act of finding previously placed items. We made the following hypotheses for both cases:

H1 Total time to search and acquire a target with the coupled is shorter than that with the decoupled.

H2 Accuracy of acquiring a target with the coupled is lower than with the decoupled as the coupled condition uses a ray-cast technique, which can be imprecise, especially for small targets.

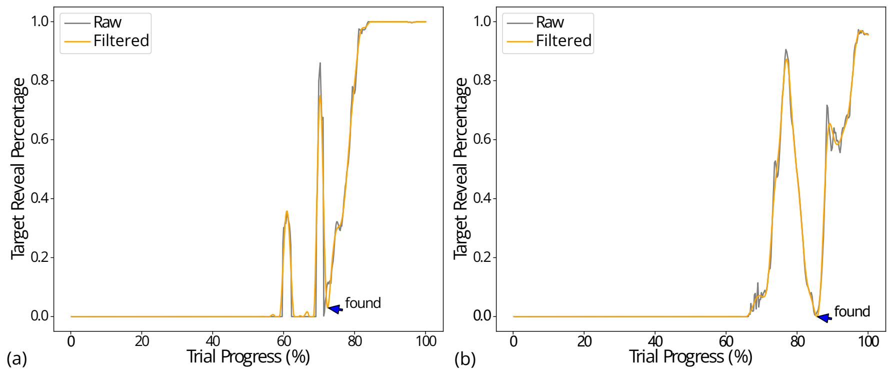


Figure 4.5: An example profile from a participant shows the reveal percentage of a target from the correct acquisition of the previous target to its final acquisition for (a) coupled and (b) decoupled conditions.

H3 Perceived physical loads and efforts with the coupled techniques is lower than the decoupled technique thanks to the single hand usage.

4.4 EXPERIMENT SETUP

4.4.1 *Task*

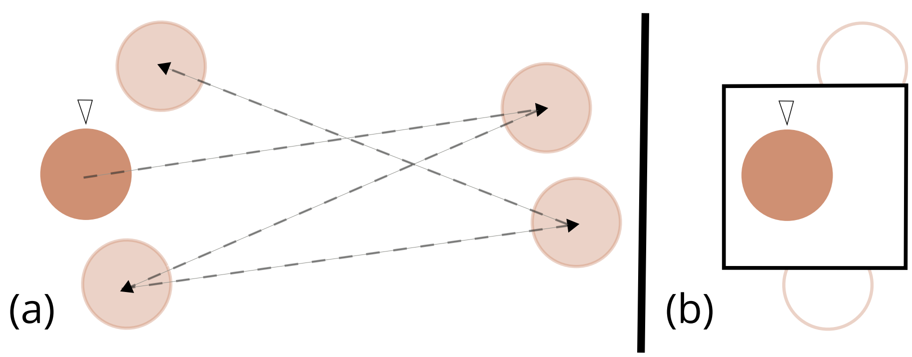


Figure 4.6: a) An example of target sequence in a set; b) only one target is presented and can be revealed by the peephole at a time on the surface.

We designed a 2D pointing task based on previous 1D dynamic peephole pointing studies [47, 76, 120, 216]. We created a surface based on the physical desk, measured approximately in 86cm × 46cm. Rather than using vertical rectangles that occupy the entire height of the surface [47, 76, 120, 216], we placed circular targets on the left and right sides of the surface to prevent participants from simply moving left or right to find targets. We aimed to simulate real-world 2D target acquisition scenarios in which participants know roughly the direction and distance of targets to search and acquire.

Peephole size was also excluded as an independent variable due to time constraints in a VR study [70, 211]. Previous studies [47, 120, 185] showed that peephole size affects acquisition performance, but larger sizes do not always improve results [47, 185]. Given the lack of consensus on the optimal peephole size for 2D tasks, we chose a 12cm $\times$ 12cm peephole as a sweet spot based on empirical findings from Rädle et al.'s work [185]. This size represents a balance between performance and preference for both experiments.

These circular targets were generated on either the left or right side of the virtual surfaces. The first trial was used to calibrate the starting position. In this trial, the circle appeared randomly on the left or right side with an inverted pyramid floating above it to indicate its position and relative direction to the next trial. Correctly selecting the circle moves the participant to the next trial, where the next circle appeared randomly on the opposite side of the previous target. Targets were distributed around two edges of an area restricted by a controlled amplitude, as shown in Figure 4.6. Participants were reminded that targets would alternate between the left and right sides when the experiment started. This design simulates our scenario where users know where to search within a certain area. Participants were instructed to search for and select the circular targets on the table as quickly and accurately as possible. Different notification sounds indicated correct and incorrect selections. If any selection was outside the bounds of a circular target, the trial was marked as an error. The error rate was displayed at the end of each block to remind participants to perform the task as accurately and quickly as possible.

4.4.2 *Procedure*

Before informed consent was obtained, participants were asked to read an information letter in which they were warned about potential motion sickness as a result of wearing HMDs and that they could stop the experiment at any time without penalty if they felt uncomfortable or simply did not wish to continue. Then they were asked to respond to a demographic questionnaire (requesting their gender, age, handedness, and VR and peephole interaction experience) and invited to sign the informed consent form.

Participants were then asked to wear the HMD, and recenter its viewport after starting the experiment system. They specified their handedness in the system, and performed a calibration procedure to match the physical and virtual environment. They then performed the fingertip calibration, and touched to select example targets on the virtual table to start the experiment sessions. After the experiment started, participants were presented with a virtual desk, a floating virtual lamp above the desk and a peephole display window on the table. The participants were guided through verbal instructions to manipulate the lamp and peephole and perform conditions before acquiring the very first target. They started to complete the task with the given condition, and were reminded of the remaining blocks and trials by a notification panel placed at a distance in the virtual environment. When participants finished a block, they had to take a break of at least 5 seconds. When participants completed a condition, they were asked to remove the HMD, complete a Task

Load Index questionnaire, and take an unlimited time break. They resumed the experiment session when they were ready. When participants completed both conditions, the experiment ended, and participants were asked for their preferences and feedback on each condition.

4.4.3 *Participants*

We recruited 16 right-handed participants, aged 20 to 36 (M = 26.2, SD = 3.10), of which 3 were female. Participants were recruited by word of mouth at a local institution. Among all participants, 14 had VR experience before, and 8 of them had more than 6 months of experience; 10 of them had experience with peephole interactions (*e.g.* Pokémon GO, and map navigation). Participants all used the non-dominant hand to grab the controller, and manipulate the virtual lamp. The experiment took around 30 to 40 minutes.

4.4.4 *Design*

A repeated measure within-participant factorial design was used. The independent variables were CONDITION (COUPLED, DECOUPLED), BLOCK (1-5), WIDTH (4cm, 8cm) and AMPLITUDE (32cm, 48cm, 64cm). The order of CONDITION was counter-balanced using a Latin square. In each block, participants performed 6 sets of 5 trials, with each set formed by a complete combination of AMPLITUDE and WIDTH. The sets were arranged in ascending order of ID and the first trial was used to calibrate the starting position in a set.

In summary, each participant completed 2 CONDITION × 5 BLOCK × 2 WIDTH × 3 AMPLITUDE × 5 Trials, i.e. 300 trials. This resulted in a dataset containing 4800 trials for our 16 participants.

4.4.5 *Apparatus*

The system was implemented in Unity 2022.3.11f1 with the Oculus Integration v57.0. The direct touch method was implemented with a Minisense 100 vibration sensor [157] connected to an Arduino MKR Wi-Fi 1010 board [10] and tested with an android smartphone and a tablet with OptiTrack systems. The experiment was conducted on the Oculus Quest 2 (tracking frequency is 72 Hz) with Oculus left- and right-hand controllers.

4.5 RESULTS

We removed the first trial from each set for position calibration purpose, followed by the elimination of outliers that exhibited total time deviating more than three standard deviations from the mean on the non-erroneous trials. This process yielded a dataset comprising 3774 trials (98.3%) for analysis.

To address the non-normal distribution of our collected samples, we employed Friedman tests (α= 0.05) with Wilcoxon signed-rank test as post-hoc analysis to examine error rates and subjective feedback. We applied a Box-Cox transformation on our data and conducted repeated-measure ANOVA (α= 0.05) for time-related metrics. When sphericity was violated, we employed the Greenhouse-Geisser correction to adjust degrees of freedom. For post-hoc analysis, we performed pairwise t-tests with Bonferroni corrections when significant effects were detected. Additionally, we modelled our time-related metrics using Cao, Li, and Balakrishnan's model [48] and Huber, Steimle, and Mühlhäuser's model [107] to explore how COUPLED and DECOUPLED methods impact target acquisition in a spatial peephole interface. Error bars in charts were 95% confidence intervals (bootstrapped with 10,000 re-samples).

4.5.1 *Learning Effect*

We found a learning effect through the analysis of both error rate and total time and removed the Block 1 in our remaining analysis to keep a consistent and stable performance analysis. The Friedman test reported a significant main effect of BLOCK on error rate ($\chi^2_{(4)}$=14.53, p<0.01) and the post-hoc analysis showed a significant difference between Block 1 (M = 5.3%) and Block 5 (M = 1.9%, p<0.05). As for total time, a two-way RM-ANOVA (α=0.05) with BLOCK × CONDITION on total time also only reported a significant effect of BLOCK ($F_{4,60}$=16.07, p<0.001, η^2=0.10). Pairwise comparisons revealed significant differences between Block 1 (M = 3.81s) and the other four blocks (p<0.01). We found no significant effect of CONDITION or significant interaction effect between BLOCK and CONDITION.

4.5.2 *Error Rate*

The mean error rate of COUPLED and DECOUPLED were 4.0% and 2.0% respectively. The Friedman test did not reveal a significant effect of CONDITION on error rate, but showed significant effects of WIDTH ($\chi^2_{(1)}$=10.29, p<0.005) and AMPLITUDE ($\chi^2_{(2)}$=6.90, p<0.05). Pairwise comparisons showed that large targets (M = 1.5%) caused significantly fewer errors (p<0.01) than small ones (M = 4.5%). No significant effect of AMPLITUDE was found.

4.5.3 *Time*

Total Time

Total time for DECOUPLED (M = 3.13s) is not significantly different from COUPLED (M = 2.95s). We found however a significant effect of AMPLITUDE ($F_{2,30}$=5.77, p=0.008, η^2=0.01) and WIDTH ($F_{1,15}$=60.52, p<0.001, η^2=0.06). Total time for selecting large targets (M = 2.82s) was significantly faster (p<0.001) than that of small targets (M = 3.26s), but no significant difference was found between the amplitudes. We also found an interaction effect of CONDITION and AMPLITUDE ($F_{2,30}$=3.50, p=0.043, η^2<0.01). However, post hoc analysis did not reveal significant differences.

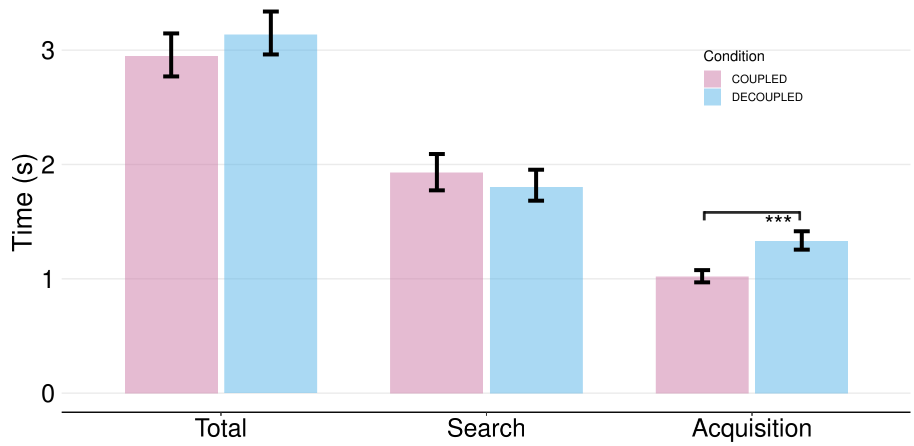


Figure 4.7: Completion, Search and Acquisition Time(s) for CONDITION.

Search Time

Search time for DECOUPLED (M = 1.80s) and COUPLED (M = 1.93s) were not found significantly different. We found a significant effect of AMPLITUDE ($F_{2,30}$=5.95, p=0.007, η^2=0.02) and WIDTH ($F_{1,15}$=8.40, p=0.011, η^2=0.01) on the search time. However, post-hoc analysis did not reveal significant difference between the two widths or among the three amplitudes. We also found a significant interaction effect of CONDITION and AMPLITUDE ($F_{2,30}$=4.18, p=0.025, η^2=0.01), but post-hoc analysis did not reveal significant differences among the combination of levels.

Acquisition Time

The acquisition time of DECOUPLED (M = 1.33s) was significantly greater ($p<0.001$) than that of COUPLED (M = 1.02s). We found a significant effect of CONDITION ($F_{1,15}$=22.60, $p<0.001$, η^2=0.21) and WIDTH ($F_{1,15}$=244.63, $p<0.001$, η^2=0.22) on the acquisition time. Participants spent significantly less time ($p<0.001$) to acquire a large target (M = 1.03s) than a small target (M = 1.33s). We also found an interaction effect of CONDITION and WIDTH ($F_{1,15}$=17.83, $p<0.001$, η^2=0.01). The post-hoc results showed that there were significant differences between all combinations of CONDITIONS and WIDTHS except for DECOUPLED method at 8cm (M = 1.19s) and COUPLED method at 4cm (M = 1.18s). For each WIDTH, COUPLED was faster than DECOUPLED.

Time Analysis using Velocity Profiles and Target Reveal Percentage

We analyzed peephole velocity and target reveal percentage to understand the time differences between COUPLED and DECOUPLED conditions. We used the search time of each trial to identify the search phase and acquisition phase, and computed the velocity based on the peephole's 3D position and corresponding timestamps. The velocity and target reveal percentage for each

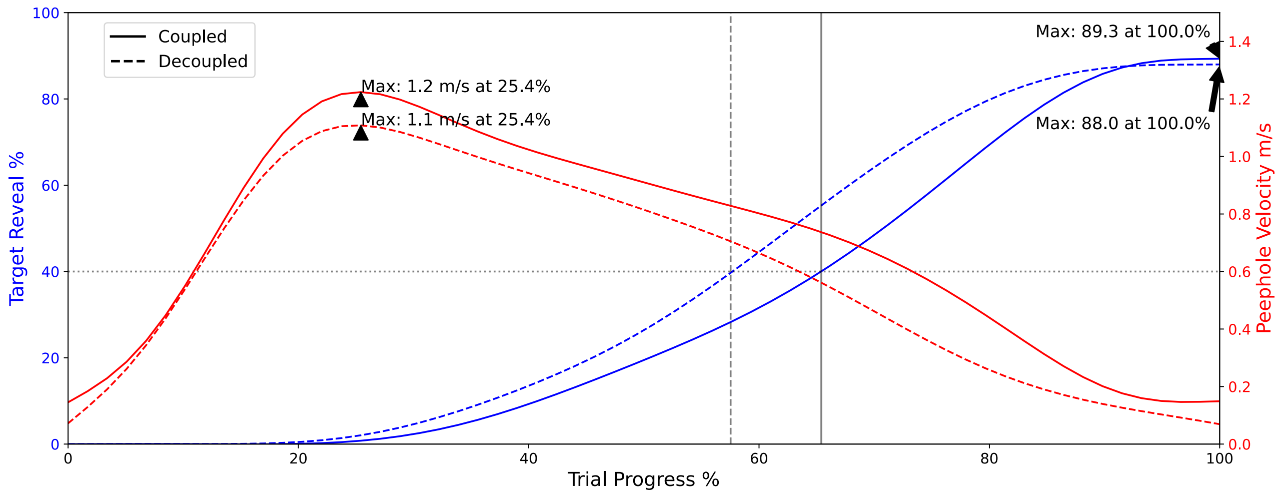


Figure 4.8: Overall peephole velocity and target reveal percentage averaged over all trials. Vertical lines are computed by mean search time and total time, and mark the separation between search and acquisition phases. The horizontal line marks target reveal at 40%.

trial were then interpolated to create a time-equidistant profile at 60Hz, and smoothed with a zero-phase filter. For each CONDITION, these values were aggregated by computing the mean for each normalized time interval. Figure 4.8 illustrates the average peephole velocity and target reveal percentage over all trials.

We observed that with the same searching technique, COUPLED exhibited trends of higher peephole velocity and lower target reveal percentage compared to DECOUPLED. Given this observation, we computed the raw peak velocity (PkV) for each trial, and performed an ANOVA analysis as previous steps. PkV of COUPLED (M = 4.10m/s) was significantly faster than DECOUPLED (M = 3.72m/s), with a main effect of CONDITION ($F_{1,15}$=17.75, p<0.001, η^2=0.21). While we found a main effect of WIDTH ($F_{1,15}$=130.15, p<0.001, η^2=0.20), post-hoc analysis did not reveal significant differences. We did not find interaction effects of CONDITION with WIDTH or AMPLITUDE.

4.5.4 *Modelling*

Our data did not fit well with CLB's model [47], where A is amplitude, W is width, and S is peephole width:

$$T = a + b\left(n\,log_2\left(\frac{A}{S}+1\right) + (1-n)\,log_2\left(\frac{A}{W}+1\right)\right)$$

or HSM's model [107], where L is surface length:

$$T = a + \frac{b}{S} tan\left(\frac{A}{L}\pi + \frac{\pi}{2}\right)$$

To perform the analysis, we compensated for the non-normal distribution of total time by first aggregating the mean for each set of AMPLITUDE and WIDTH

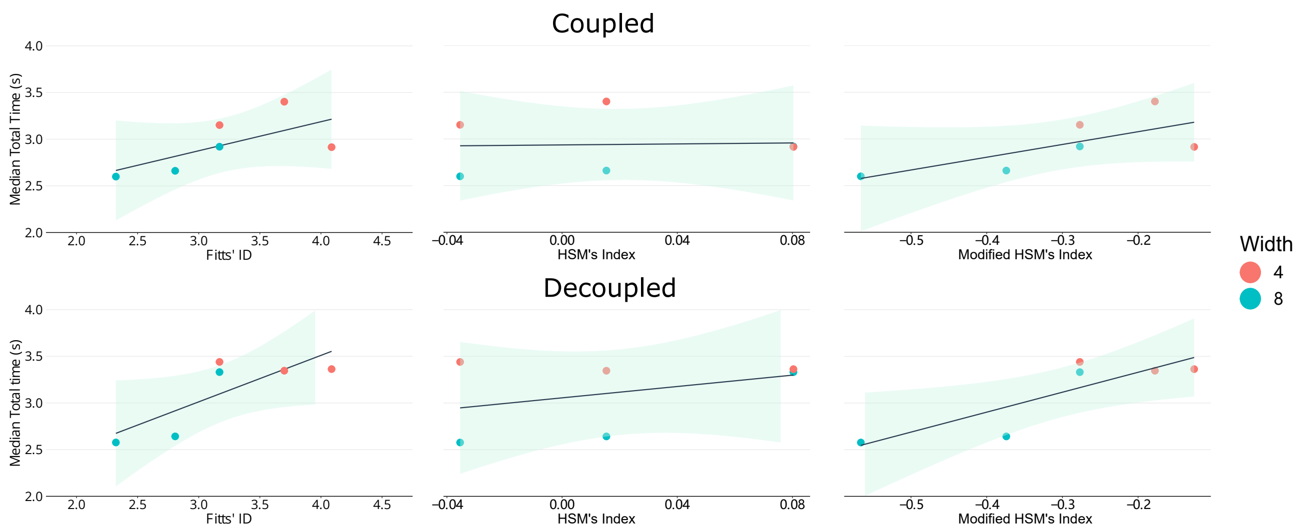


Figure 4.9: Modelling results: COUPLED (top) and DECOUPLED (bottom). Left to right are Fitts' Law modelling, HSM's modelling, and modified HSM's modelling.

per BLOCK per CONDITION and aggregating the median. The results provided a value of n=0 for both COUPLED and DECOUPLED conditions in CLB's model, leading to the results of a standard Fitts' law modelling. The corresponding R^2 for COUPLED and DECOUPLED conditions for Fitts' law are 0.42 and 0.62 respectively. As for HSM's model, it gave us R^2=0 for COUPLED and R^2=0.16 for DECOUPLED. However, as HSM's model did not consider the impact of targets' width, we introduced a modified model based on previous results [47, 120, 185, 196] where target width and amplitude are inversely correlated for movement in dynamic peephole interaction:

$$T = a + \frac{b}{S} tan\left(\frac{A}{L\mathbf{W}}\pi + \frac{\pi}{2}\right)$$

This model gave us slightly better results with R^2=0.51 for COUPLED and R^2=0.72 for DECOUPLED. Our modelling results were presented in Figure 4.9.

4.5.5 *Task-Loads, Preferences and Interaction Strategies*

Overall, the Wilcoxon signed rank test only reported significant effects of CONDITION on TEMPORAL ($\chi^2_{(1)}$=32.5, $p<0.05$). Six participants preferred COUPLED; five preferred DECOUPLED; five believed that there are pros and cons for each method, and they should be used depending on the context. The main advantages for COUPLED included its convenience, ease of use with a single hand and reduced physical demand. As commented by P1, *"it (Coupled) is more convenient and comfortable and there is no need to use right hand to select, which is tiring."* However, COUPLED was reported with issues with precision, particularly when selecting smaller targets, *"The 2nd (coupled) is way way way easier for the bigger targets and there was a bit of jitter for selection."* [P5]. In contrast, DECOUPLED was preferred due to its increased precision and a more

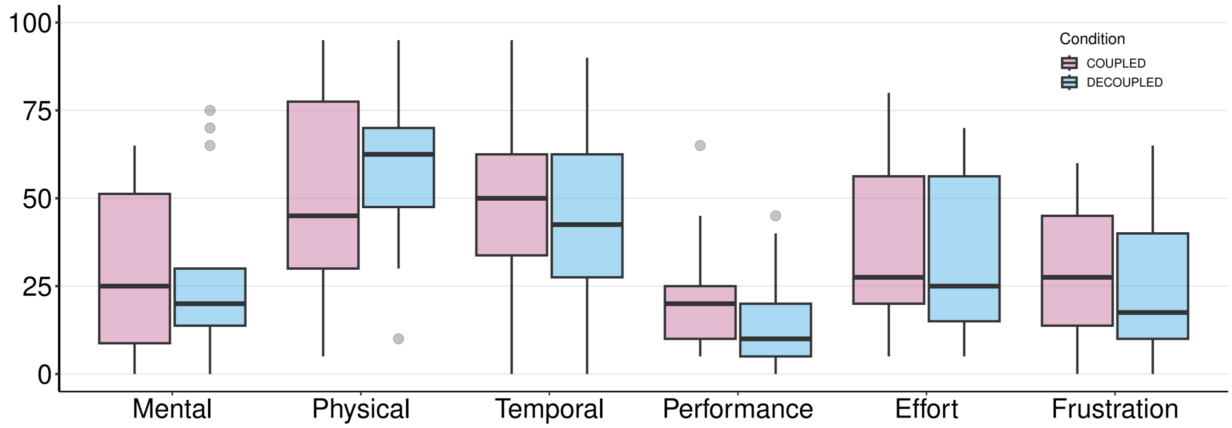


Figure 4.10: TLX Scores by CONDITION. Lower scores are better.

engaging experience. P6 highlighted this as *"I guess I did not have to focus on the left arm ... and I could do the precise selection with my dominant hand, "* and *"I don't like do thing with one hand and another hand doing nothing."* Nearly a third of the participants expressed mixed preferences, noting their preferred method depended on the specific task at hand, which was well-stated by P9 as *"It depends on what I am doing. Having everything with one hand is a bit tiring, and I am way more precise with finger, but I could be a bit slower."*

We noted that participants had different strategies for COUPLED and DECOUPLED conditions. With COUPLED condition, participants focused on arm stability where they either rested their arms on the table or tightly next to their body and mainly rotated the grabbed controller using their wrist to acquire targets. In contrast, the DECOUPLED condition involved more and larger body movements and coordination, where participants lifted the controller in the air and moved back and forth in front of the desk for more dynamic and exploratory movements. Besides, participants were observed to initially use translation movements and then adapted to rotational movements as needed. This could possibly be due to fatigue after long-duration of body movements.

4.6 DISCUSSION

We did not confirm hypotheses expecting differences between coupled and decoupled target acquisition (Table 4.3). Notably, error rates were low in all conditions and the total times did not differ significantly. Overall, these suggest that peephole dynamic pointing is reasonable for desktop surface environments, whether using coupled or decoupled target acquisition.

Look closely, we cannot provide a one-size-fits-all answer to the question *whether target acquisition in a dynamic peephole on augmented surfaces should be coupled or decoupled*. We did not find significant differences in total time and accuracy between COUPLED and DECOUPLED. More specifically, we did not find a significant difference between COUPLED and DECOUPLED in search time,

Hypothesis	Experiment Results
H1: Total time to search and acquire a target with the coupled is shorter than that with the decoupled.	**Not Supported**: COUPLED did not have significantly shorter total time than DECOUPLED
H2: Accuracy of acquiring a target with the coupled is lower than with the decoupled.	**Not Supported**: COUPLED did not have significantly lower accuracy than and DECOUPLED.
H3: Perceived physical loads and efforts with the coupled techniques is lower than the decoupled.	**Not Supported**: COUPLED did not have significantly lower physical loads and efforts than DECOUPLED.

Table 4.3: How our results relate to our research hypotheses introduced in Section 4.3.5.

but we did in acquisition time. For the accuracy, it is possibly related to the interaction techniques we used, as we found similar accuracy performances for both techniques as in Ens, Ahlström, and Irani's work [76]. Participants commented that it was difficult to select small 4 cm targets with COUPLED due to hand jitter. This can be attributed to the ray casting technique, known to make small targets difficult to select [18], and also to the potential Heisenberg effect when pressing the button [255]. In fact, previous work reported mixed answers to this question, either reporting the coupled condition showed obvious advantages in a desktop setting [47], or finding the decoupled methods more efficient [216]. Given these conflicting results, it is not so surprising that no clear difference was found.

While COUPLED and DECOUPLED had a similar search time, they led to different velocity profiles in the search phase: peephole moved significantly faster but revealed targets later with COUPLED, compared to DECOUPLED. The participants commented that they could easily interact using COUPLED and that it required less physical effort. They were also observed to have minimal body movement and mainly perform wrist rotation with COUPLED and more diverse and large body movements with DECOUPLED. However, the analysis of the NASA TLX results did not show significant differences among factors, suggesting that the physical demand is in fact similar for the two techniques. These quantitative results, together with our observations on participants' physical movements, suggested that participants changed how they interacted and manipulated the same peephole device when they planned to acquire targets in the dynamic peephole in different ways with different control conditions.

In a standard target acquisition task, targets are visible and selection time depends on the distance to it. As the direction of the target is known in a 1D dynamic peephole target acquisition, this can explain why the standard and CLB's models still work. However, the nature of the 2D peephole task required participants to consider both direction and distance, and not only to move in a unidirectional way until reaching it. This was confirmed by the modelling where our data poorly fitted the CLB's model [47], unlike other works [76, 120, 216]. This leads to a consideration of the impacts of the potential direction in such HSM modelling [107]. While our data did not fit the HSM model, we considered the integration of the target width which improved the results, especially for DECOUPLED.

4.6.1 *Design Implications*

Based on the findings of our studies, we draw a set of design implications for dynamic peephole interaction on augmented desk surfaces with coupled and decoupled techniques.

Dual Condition Offerings

Given the different performance characteristics of COUPLED and DECOUPLED, a system should consider allowing both acquisition techniques when designing dynamic peephole interfaces on augmented surfaces. Since the COUPLED and DECOUPLED techniques rely on different input devices or interaction techniques on augmented surfaces, both modes should remain accessible at all times. For instance, we could use COUPLED for tasks requiring rapid acquisition with larger interaction area, such as picking up a window, or when a target is simply out of arm reach, and leave DECOUPLED for other tasks, such as pressing a button within that window, or when targets are easily reachable.

Designing Novel Interactions

When designing interactions based on coupled and decoupled conditions for dynamic peephole target acquisition, one should consider their impacts on the search and acquisition phases. While the search time could be similar, techniques based on coupled conditions may cause potentially faster peephole movement, but reveal targets later, while techniques based on decoupled conditions may lead to slower peephole movement, requiring better coordination, but reveal targets earlier. Techniques should consider which phase it optimizes to improve the overall interaction phase.

Augment Peephole with Visual Aids

Our results revealed that 2D target acquisition in dynamic peephole leads to a more difficult search and acquisition process compared to a 1D task, contrasting with previous studies [47, 76, 120, 216]. We did not use visual cues to guide users, or indicate the direction and distance to target, because we wanted to simulate the scenario in which participants had to search in a specific area. However, for more complicated and hard-to-reach surfaces [56], such helps could indicate where available targets are, and reduce target search time.

4.6.2 *Limitation and Future Work*

Peephole Size and Prior Knowledge as Factors?

Our experiment did not include the study of the peephole size on performance due to the duration of the VR study [70, 211], and we defined the size of the peephole based on previous results [185]. Instead our experiment controlled the regions where targets appear, determining the amount of prior knowledge participants had about the direction and distance of the targets. Our work thus establishes a baseline for 2D dynamic peephole pointing in absence of visual cues. Future work should investigate the impact of peephole size, prior knowledge, and visual cues in a more systematic way.

Handedness for Dynamic Peephole

Participants commented that they might perform better with their dominant hand for COUPLED. In our experiment, we initially wanted to reduce the confounding effect of handedness for manipulating the peephole, and we found that this concern may not be the case since both hands and arms were lifted and used for DECOUPLED. Besides, as P6 explained, *"I would still have the gorilla effect with the right hand with the coupled technique, and I do not think there is much difference."* Future work can explore the impact of handedness on dynamic peephole target acquisition.

Diverse Tasks and Environments

Our studies primarily focused on target acquisition within dynamic peephole interactions. However, augmented surfaces support a wide range of tasks that remain unexplored, such as tracing and docking. Future work could investigate the impact of control mechanisms on different tasks, especially considering the single hand usage for the coupled technique across different surfaces. Our studies conducted experiments in a VR environment to avoid confounding factors. While it is tempting to assume that our results can be translated into mixed reality environments, this assumption needs to be further assessed.

4.7 CONCLUSION

In this chapter, we investigate the performance and preferences of coupled versus decoupled methods for target acquisition in dynamic peephole interactions within augmented desktop environments (RQ4). This work extends previous research by examining target acquisition in 2D dynamic peephole interaction, and highlighting the limitations of existing models. Our findings reveal subtle differences in performance metrics, with coupled method showing significantly faster acquisition time. Coupled method is preferred for its convenience, and reduced physical load. Meanwhile, decoupled method is preferred for its accuracy. More interestingly, we also find characteristic differences when performing the same search task with two control methods. Participants tended to move the peephole faster with the coupled method but reveal targets later. In contrast, they moved the peephole slower but reveal targets earlier with the decoupled method. In addition, we also observed that participants mainly rest their arms on the desk or close to body, and rotate their wrists for aiming and acquiring targets with the coupled condition, But, with the decoupled method, they tended to lift their arms, and move their body for more exploratory movements. Based on these findings, we purpose different design guidelines for the use of coupled and decoupled methods for dynamic peephole interaction, and investigate a more general form of this phenomenon in the following chapter.

5 EXPLORING THE EFFECTS OF INTENDED USE ON TARGETING IN VIRTUAL REALITY

Despite employing the same searching technique for dynamic peephole interaction, coupled and decoupled methods exhibit distinct patterns in velocities and target reveal percentages during the search phase, and in overall interaction strategies. These differences may be influenced by the intended use of the device for an interaction technique. We extend this observation in a more general context by investigating whether what users intend to do with a virtual target impacts how they plan and perform the initial target acquisition.

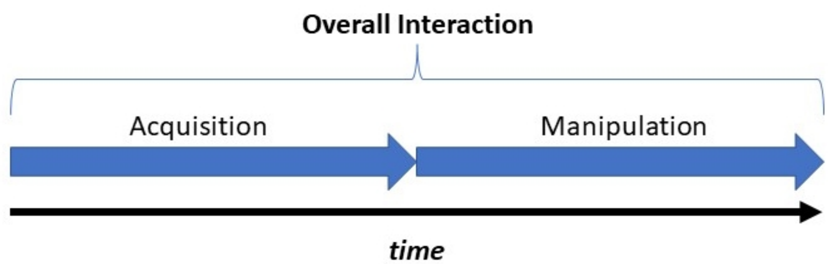


Figure 5.1: The overall interaction with an object includes two sequential steps: acquiring the object – time characterized by Fitts' Law or a variant – and manipulating the object – time varies depending on the complexity and nature of the manipulation.

5.1 INTRODUCTION

Target acquisition is one of the most common actions performed in an interactive system regardless of whether the computing platform is a desktop computer [92, 146], a touch-based device [90], a large physical display [214] or an Augmented Reality environment [57].

Fitts' Law [146] is the most commonly used model to describe the movement time taken to acquire a target. Movement time (*MT*) for a one-dimensional pointing task is described by a linear function of the index of difficulty (*ID*) of the pointing task:

$$MT = a + b \cdot ID \quad (5.1)$$

where a, b are empirically determined regression coefficients and ID is a logarithmic term of target amplitude (A) and target width (W) [145]:

$$ID = log_2\left(\frac{A}{W} + 1\right) \quad (5.2)$$

However, target amplitude and width are not the only factors impacting acquisition performance [265]. One factor that may impact target acquisition time is the *intended use* of the target. By *intended use*, we refer to manipulations that a user intends to do or will do with the target they acquire. For example, in virtual environments, beyond simply acquiring an object, users can scale objects and move them around in the space [256], rotate objects to reveal the occluded view [59] or manipulate objects' motion to simulate physical phenomena [11]. In order to perform any of these intended uses of a target, we must first acquire it. Acting on an object in an interface is thus a compound action comprised of acquiring it and performing the action or manipulation of it (See Figure 5.1) and the question we pose in this paper is whether the manipulation that we will perform on an object in a virtual environment, our intended use, impacts the performance of acquiring the target.

In 2D computer interfaces, research has shown that target acquisition time may vary depending on whether the user wishes to simply acquire a target (e.g. targeting), move it (e.g. dragging or docking), or throw it (e.g. flicking) – independent of the time taken by the subsequent task [149, 199]. It is always tempting to assume that results in 2D can be directly applied to 3D, particularly to Virtual Reality (VR), but we hesitate to make such assumption for several reasons. For instance, the controller in VR is an absolute input while the mouse is a relative input, leading to different acquisition and manipulation behaviors, and some tasks more frequent in 3D were not investigated in 2D (e.g. reorienting a target). Besides, to the best of our knowledge no similar analysis has been performed in the increasingly common context of direct target acquisition in VR. As we incorporate VR into both work [69, 136] and entertainment [212], understanding the impact of various independent variables on performance measures for VR-based selection tasks remains an area of active research interest (see [25] for a review).

To explore the impact of intended use on targeting in virtual environments, more precisely in the context of proximal interaction using the virtual hand metaphor, we examine five common manipulations in various VR systems, such as games [98], social [154] and educational applications [91] – TARGETING (selecting a target), DUALTARGETING (selecting one then another target), THROWING (acquiring and pushing a target), DOCKING (acquiring and placing a target) and REORIENTING (rotating a target about its axes) – and measure how long it takes to acquire the target to be manipulated. We find that simple targeting exhibits the fastest acquisition times for that target, reorienting the slowest acquisition times, and other intended target manipulations result in acquisition times between these extremes. Given the initial finding that the intended use of the target impacts time, we probe additional characteristics of movement toward a target with the goal of understanding how and why prior movement time is impacted by the subsequent intended use of the target. Movement profiles highlight both characteristics and differences in peak speed, time to peak speed, movement prior to selection and selection speed across different intended uses.

This paper is organized as follows. After presenting the related work on target acquisition modelling in both 2D and 3D, we describe and explain the experiment setup and task design of our study. Then, we detail our analysis from three aspects: selection time, motion kinematics and Fitts' law modelling. Finally, we present our findings and discuss their implications.

In summary, this is the first work to explore the impact of intended use on targeting in VR and our contributions include:

- Investigating impacts of intended use on target acquisition in VR and summarizing characteristics of each use.
- Presenting design implications for interfaces and interactions in VR.

5.2 PROJECT-SPECIFIC BACKGROUND

Fitts' Law [82, 146] is probably the most well-examined relationship in human-computer interaction (HCI) research. While Fitts' Law was originally formulated in terms of a 1D pointing task, researchers in HCI have long recognized that understanding the cost of target acquisition in graphical interfaces is useful, as it allows us to characterize the relative efficiency of different arrangements of interfaces. As a result, researchers have proposed a number of extensions to Fitts' Law to model 2D targeting [4], 3D targeting [225], gaze-based targeting [268], foot-based movement [96, 233], among others. Researchers have also explored modelling error in Fitts' Law [252], generalized Fitts' Law to incorporate steering tasks [3, 166], and leveraged the fundamental components of Fitts' Law to design a host of pointing facilitation techniques (see [15] for a review).

The goal of research into Fitts' Law is to understand and improve throughput [265] in the use of interactive systems. If we characterize the temporal cost and error rate of individual interactions (e.g. selecting a target, manipulating that target, keyboarding, homing, etc.), then the overall temporal cost and error rate of a compound interaction is the sum of these basic interactions [49, 71]. By improving the speed of target acquisition via new target acquisition techniques [15] and the speed of interactions via new interaction techniques [25], each of these subtasks becomes more efficient, increasing the overall efficiency of the interaction. Essentially, the assumption is that each individual interaction can be independently optimized. This is equally true in VR research: Bergström et al. [25], analyzing 20 years of VR research, note that studies with selection tasks measure the time a participant needs to select the next target (occasionally with additional measures of error and throughput for the selection task), while studies with manipulation tasks measure the time a participant manipulates virtual objects.

While, to the best of our knowledge, the independence of basic interactions has never been evaluated in VR, we have reason to believe that this assumption is questionable in traditional two-dimensional computer interfaces. Mandryk and Lough [149] examined how the *intended use* of a target impacts the time it takes to acquire the target. Mandryk and Lough note that, in real-world interfaces, the user acquires a target with a specific goal in mind. Perhaps they wish to click the target (targeting). Perhaps the target activates a secondary set of widgets, and they need to then click on a second target (i.e. dual targeting). Perhaps they wish to move the target in some way, e.g. to re-position it imprecisely (flicking or dragging) or to re-position it precisely at a new location (docking). Mandryk and Lough found that, if the intended use was targeting or dual targeting, participants acquired the target significantly faster than if the intended use was flicking or dragging. They also noted a difference in acceleration and deceleration during the selection of the target to be manipulated: if the participants intended to flick or dock the target, then the selection movement toward that target exhibited a higher peak speed and a longer deceleration phase. Follow-on work by Ruiz and Lank [199] replicated these results and, via a more complete analysis of movement profiles, analyzed their potential impact on kinematics-linked endpoint prediction [134]. While Ruiz and Lank note that the impact on kinematics-linked modelling was likely not a concern, in both cases results implied that the potential benefits of new interaction techniques observed during manipulation may not be realized if they result in a corresponding increase in acquisition time for the target to be manipulated.

There is also a significant possibility that virtual reality manipulations will differ from both real-world and two-dimensional interface manipulations. Considering real-world manipulations, the field of psychology has actively studied the act of reaching and grasping for many years [198]. Factors, including perception of the object to be grasped [87], the manipulations to be performed on the object [209], and tactile feedback during the act of grasping [113] impact both trajectories toward an object and the positioning and speed associated with the grasping of an object. However, there exists a disconnect between real world affordances and perceived affordances [170] and an absence of the physiological interactions between hand and object, which may impact behavior. Furthermore, while it is always tempting to assume that previous results in 2D [149, 199] can be directly applied to 3D, especially to VR, past research indicates that this may not be true. As one example of this, Cockburn and Mckenzie [63] found that user performance deteriorated for a locate-and-point task when transforming from a 2D interface to a 3D interface. Furthermore, movement planning, whether in real world or in computer interfaces, requires trajectory planning [134, 198]; in immersive VR environments, depth has been found to greatly impact both perceived width and distance [12, 58], a factor that may significantly alter the trajectory, kinematics and the impact of intended use.

5.3 methodology

We conducted a controlled experiment to investigate the effect of the intended use of a target on the time taken to acquire it in VR. We focused on interaction at arms' length using a 1-to-1 mapping of physical controller movement to virtual controller movement, a direct 3D target acquisition technique common in VR [58, 178, 179]. Figure 5.2 depicts the timeline of a trial in this experiment.

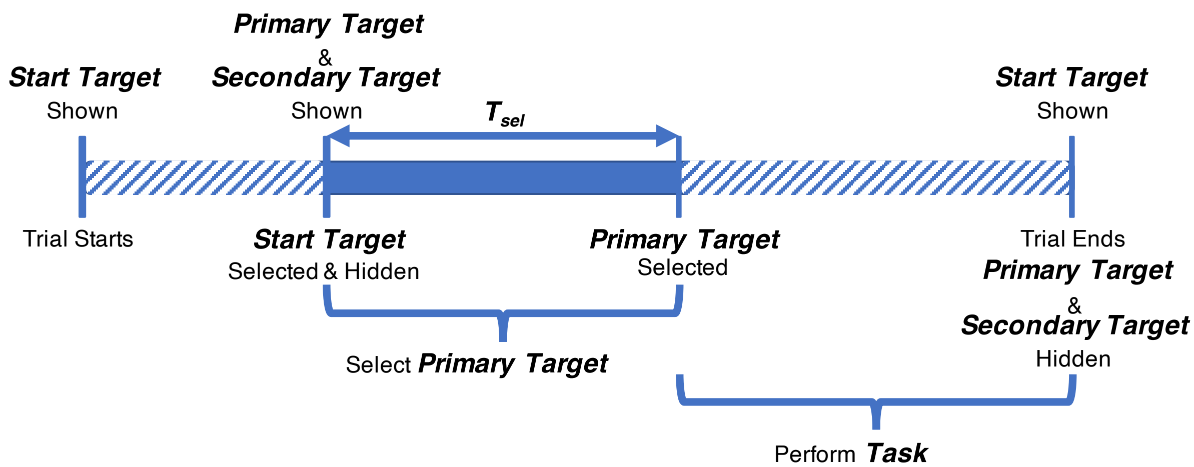


Figure 5.2: A timeline of a trial: a trial starts by showing the *Start Target*. Once it is successfully acquired, it vanishes and a *Primary Target* is shown, together with a *Secondary Target* if required by the *Task*. Participants acquire the *Primary Target* and perform the *Task*. *Start Target* is shown again when the trail ends.

5.3.1 *Apparatus and Participants*

The system was implemented in Unity 2020.3.7f1 with the Oculus Integration v29.0 and the study was conducted standalone on the Oculus Quest 2 (tracking frequency is 72 Hz) with Oculus left- and right-hand controllers.

In pilot studies we noted that the virtual controllers could occlude targets and their irregular shape caused problems of precision when selecting a target (because the exact selection location was unclear). To address this, we used a smaller controller 3D model (from GEAR VR), made the models translucent, and added a blue sphere (**cursor**) to the top side of the virtual controller indicating selection location, as shown in Figure 5.3. Although the physical and virtual controllers are not in the same shape, this design rarely affects how participants recognize them as participants only see virtual controllers rather than physical ones. In addition, participants focus on using **cursor**, rather than the model for interactions. Making the model semi-transparent also does not impact the selection performance of a virtual hand [129, 232]. Besides, **cursor** is useful for target selection in VR [17] and our design presents better visual tracking on it.

We also noted that, as participants became fatigued, they would sometimes switch hands for a brief period, interacting with their non-dominant hand which impacted performance. Therefore, alongside instructing participants to perform the study only with their dominant hand, we disabled the non-dominant hand controller at the outset of the study and invited participants to take breaks during the experiment if needed.

A total of 15 participants, aged from 22 to 33 ($M = 26.1, SD = 3.1$, 5 identified as women, all right-handed) participated in the study. 12 participants had experienced VR prior to the study; those participants primarily used VR for entertainment activities, such as playing games. The experiment took approximately 45 minutes and participants received $15 for their participation.

5.3.2 *Interaction for Target Acquisition*

Figure 5.3 illustrates the visual interaction of the target acquisition. To acquire a target, participants place the cursor (1 cm width) partially inside a target and press the selection button (in our implementation, A on the right controller) to select that target. When any part of the cursor enters a target, the target is highlighted in translucent blue, indicating that it is now selectable. A successful selection turns the cursor green. If instead the button is pressed while the cursor is outside the target, the cursor turns red and an alert sound is played to inform participants of the erroneous selection. Participants are expected to correctly acquire each target, and they can proceed after a correct selection.

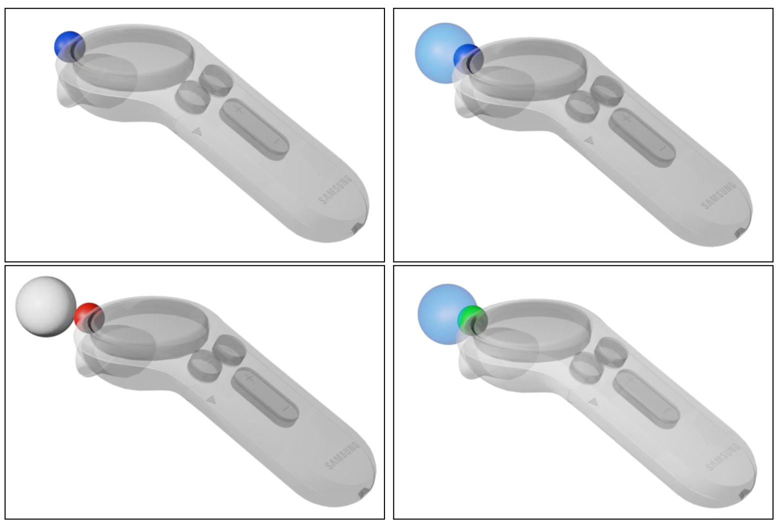


Figure 5.3: Input controller: when *Cursor* enters a target, the target turns translucent blue. Incorrect selections turn *Cursor* into red while correct selections turn it into green.

Given that the overall goal of this experiment was to measure how the intended use of a target impacts the time taken to acquire that target, the overall interaction requires acquisition of a target followed by a manipulation of that target, i.e. an intended use or *Task*. The interaction to start a *Task* proceeds as follows:

1. A participant acquires an initial target, the *Start Target*. To ensure that all targets in different tasks are noticeable and reachable within the field of view, the *Start Target* is positioned 35 cm in front of and 20 cm below the head.

2. When the participant correctly selects the *Start Target*, the *Start Target* vanishes and a *Primary Target* is displayed immediately. He/She is asked

to acquire and manipulate this *Primary Target*. Considering that the arms of healthy adults typically reach at least 60 cm from the torso [12, 58, 127, 132, 176], targets are placed within 60 cm of participants in a region surrounding the location of the *Start Target*.

3. Participants move the cursor from the *Start Target* to the *Primary Target* and perform a *Task* on the *Primary Target*.

By default, the *Start Target*'s position remains the same in the virtual world for all tasks and targets are anchored relatively to the *Start Target*'s position, which do not follow participants' movement. However, as participants could adjust their position during the experiment, we allow them to re-calibrate the *Start Target*'s position when they notice a shift in their position, before starting any new *Task*. In practice, the re-calibration was seldom used.

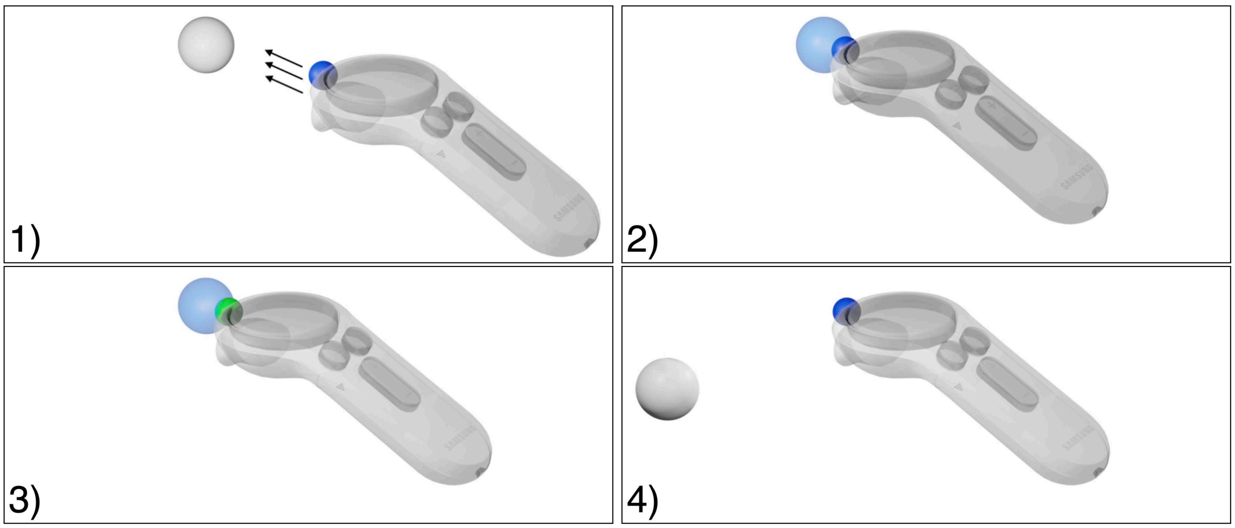


Figure 5.4: Correctly acquiring *Start Target* to calibrate starting position and reveal the task (the example shows TARGETING).

5.3.3 *Independent Variables*

Independent variables (IVs) in this study include the Index of Difficulty (*ID*) of the *Primary Target*, *Primary Direction*, *Secondary Direction* and *Task*.

ID of the acquisition action is computed using MacKenzie's formula [145] for the Amplitude **A** (distance between the *Start Target* and the *Primary Target*) of values 9 cm, 12.5 cm, or 16 cm, and the Width **W** of the *Primary Target* of values 3 cm or 5 cm). This yields six different IDs (1.49, 1.81, 2.00, 2.07, 2.37, and 2.66 bits).

Primary Direction **D** and *Secondary Direction* **D′** (Up, Down, Forward, Backward, Left or Right) represent the direction of movement from the *Start Target* to the *Primary Target* and the *Primary Target* to the *Secondary Target* respectively.

The rationale for these values is as follows. The *Primary Target* can be located in six basic directions from the *Start Target*, i.e. Up, Down, Forward, Backward, Left and Right. Given the position of the *Start Target*, six distinct directions, and the possible existence of sequential manipulation after acquiring the *Primary Target*, the maximum values of target amplitude and target width are set to 20 cm and 10 cm respectively to avoid a target appearing outside the field of view or outside the reachable workspace. These constraints motivate the above values for independent variables. With this configuration, participants observe targets that are close to them from a top view rather than a straight horizontal view, which helps to reduce the depth impact on the perceived width and addresses the occlusion issue between *Primary Target* and *Secondary Target*.

5.3.3.1 *Task*

The *Tasks* that participants were asked to complete were one of five manipulations of the primary target: TARGETING, DUALTARGETING, DOCKING, THROWING, or REORIENTING (illustrated in Figure 5.5). Detailed description of individual tasks follows.

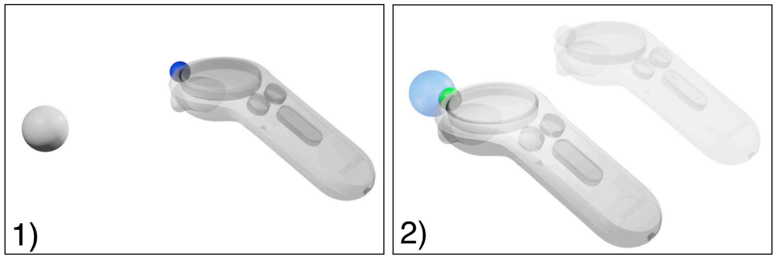
Targeting: Correctly selecting the *Start Target* of width W reveals a white sphere (*Primary Target*) of width W located at the amplitude A in direction D from the position of *Start Target*. Participants simply have to acquire the *Primary Target* by moving the controller to the target and selecting it.

DualTargeting: Correctly selecting the *Start Target* of width W reveals a white sphere (the *Primary Target*) and a red sphere (the *Secondary Target*) both of width W. The *Primary Target* is located in direction D with amplitude A from *Start Target* while *Secondary Target* is located in another direction D' (different from previous and next trials) with amplitude 9 cm from the *Primary Target* (a distance that guarantees that both targets still remain within arms' reach for the participant, see Section 5.3.3). Participants must select these two spheres sequentially: first the *Primary Target*, and then the *Secondary Target*.

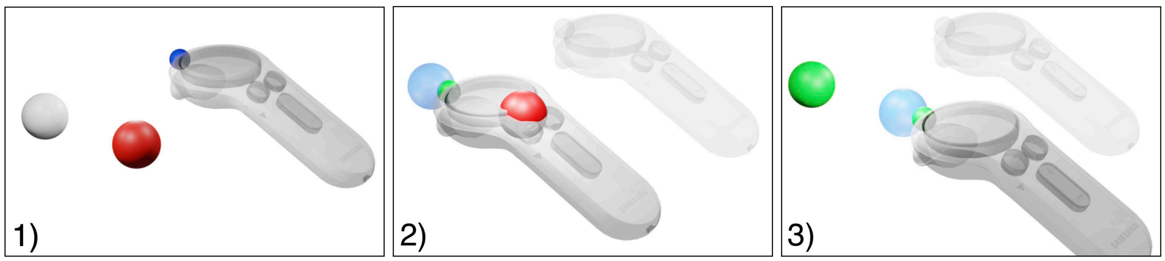
Docking: Correctly selecting the *Start Target* of width W reveals a white sphere (the *Primary Target*) of width W and amplitude A in direction D, and a semi-transparent sphere (the *Secondary Target*) 1.5 times larger than the *Primary Target* and located 9 cm away from the *Primary Target* in direction D'. The width of the *Secondary Target* reduces the required precision of the task, allowing participants to finish this task more easily. Participants are instructed to drag-and-release the *Primary Target* into *Secondary Target*.

Throwing: Correctly selecting the *Start Target* of width W reveals a white sphere (the *Primary Target*) of width W located at a position direction D from *Start Target* and at a distance of amplitude A, and a semi-transparent green wall (the *Secondary Target*) located 9 cm away from the *Primary Target* in direction D'. To reduce task difficulty (wall size being too small) and avoid visual distraction (wall size being too big), the size of this wall is set to $40\,\text{cm} \times 40\,\text{cm} \times 1\,\text{cm}$ which is over 8 times larger than the size of *Primary Target* in width. Participants are instructed to *"throw"* the *Primary Target* towards the *Secondary Target* by releasing the pressed button. The released *Primary Target* then moves in the throwing direction.

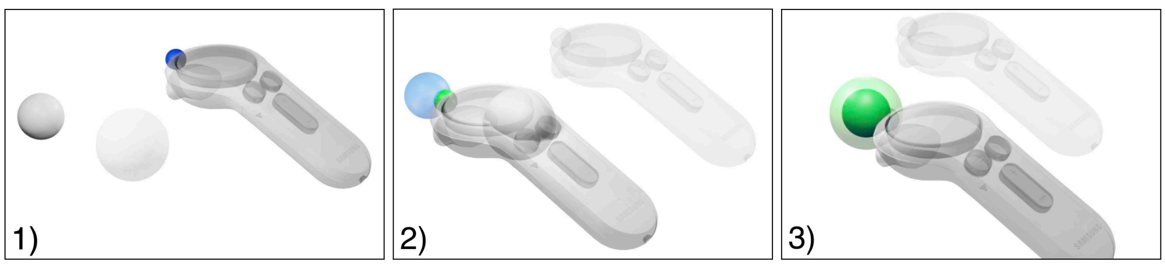
(a) TARGETING



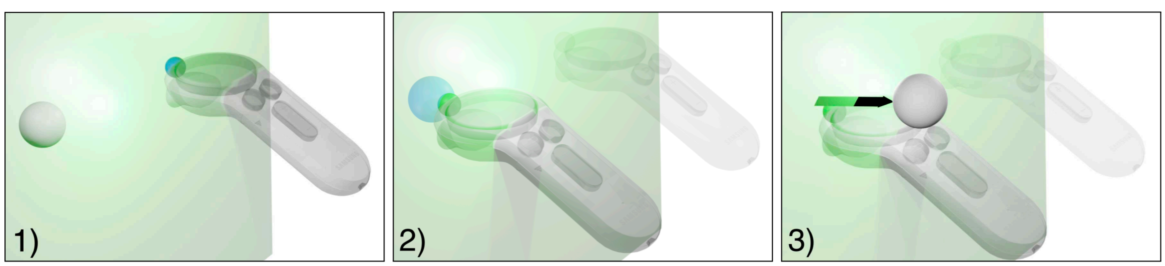
(b) DUALTARGETING



(c) DOCKING



(d) THROWING



(e) REORIENTING

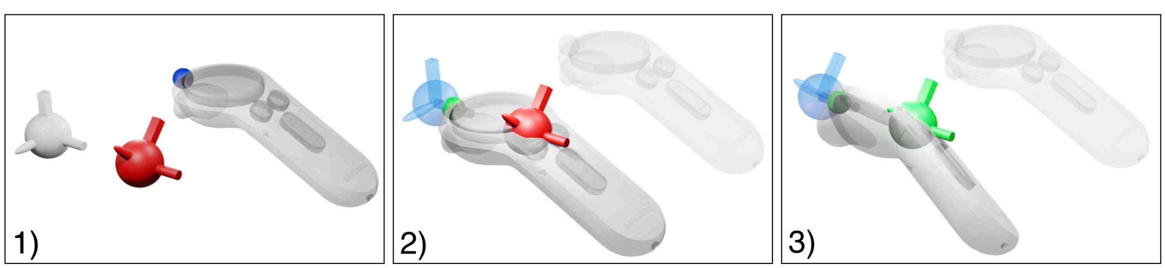


Figure 5.5: Illustration of visualization, selection and manipulation actions in each task. 1) *Primary Target (white) appears at the left of the vanished Start Target and Secondary Target appears at the back of the Primary Target. 2) Correct selection on the Primary Target. 3) Perform the manipulation actions on either Primary Target or Secondary Target based on the task.*

Reorienting: Correctly selecting the *Start Target* of width W reveals a white object (the *Primary Target*) and a red object (the *Secondary Target*) whose bodies are both spheres of width W. Different shapes (i.e. a capsule, a cylinder and a cube) are placed into both *Primary Target* and *Secondary Target* to indicate the orientation, which only serve as visual references and are not selectable. The *Primary Target* is located at amplitude A from *Start Target* in direction D and *Secondary Target* 9 cm away from *Primary Target* in direction D'. The *Secondary Target* has an orientation along its roll axis with a random angle in the range ($-\frac{\pi}{2}$ to $\frac{\pi}{2}$), values empirically obtained from pilot studies to reduce clutching and task difficulty. To move to the next trial, participants were asked to rotate the *Primary Target* such that the *Primary Target* has a similar orientation to the *Secondary Target*, that is when the angle difference in each axis is smaller than $\frac{\pi}{12}$. The color of *Secondary Target* changes to green to inform participants of the correct orientation of the *Primary Target*.

Except TARGETING, all other tasks require a manipulation sub-task on either the *Primary Target* or the *Secondary Target*. For DUALTARGETING, a second selection is required on the *Secondary Target* and other tasks require a "press and hold" behavior on the *Primary Target* for object manipulation. While behaviors in the manipulation sub-task vary across *Task*, the acquisition time of the *Primary Target* (delimited by the controller "button down" action on the primary target) is consistent across all tasks.

5.3.4 *Procedure*

Participants were recruited from our local university. To preserve social distancing requirements, the study was conducted remotely, and both the VR headset and controllers were sanitized before being delivered to participants.

Before written consent was obtained, participants were asked to read an information letter in which they were warned about potential motion sickness as a result of wearing the head-mounted display device and that they could stop the experiment at any time without penalty if they felt uncomfortable or simply did not wish to continue. They were then asked to watch an instructional video, answer a demographics' questionnaire (asking for their gender [220], age, handedness, and VR experience). They were then asked to sign the informed consent.

Participants started the experimental software by connecting with one of the researchers via video conferencing tools. The researcher verified informed consent, walked participants through disabling the non-dominant hand controller, and then guided the participant to start the experiment. In addition to researchers' verbal guidance, participants could also follow the instructions text in the system during the experiment, which presented detailed steps to guide them to control the system and finish each task.

As noted above, to evaluate the impact of intended use on target acquisition in VR, we represented different intended uses as a *Task* for the participant. The experiment consisted of repeated blocks of trials where participants had to select a target and complete a given *Task* on that target using the controller in their dominant hand.

Each trial, therefore, consisted of the following steps. First, the participant had to acquire the *Start Target* to calibrate a starting position. Correctly selecting this target would hide it, start a countdown timer, and reveal a *Primary Target* that the participant was asked to acquire and manipulate to fulfill the *Task* (Section 5.3.3.1) for the given condition. A trial ended once the *Task* was completed successfully or if the countdown timer exceeded 15 seconds. The countdown timer was implemented to avoid excessive trial completion times, i.e., to limit study duration. The experimental system then moved to the next trial and revealed the *Start Target* again (see Figure 5.4 & Figure 5.2).

Participants were instructed to complete the trials as quickly and accurately as possible while keeping an error rate below 5% (error rate was displayed to the participant at the end of each block). In case of an error, the trial was appended to the end of the current block, and participants needed to complete all trials within a block successfully before moving on to the next block, thus ensuring the same number of correct trials for all conditions. While participants were not required to repeat a block if the error rate exceeded 5%, controlling the error rate is a common practice to balance speed and accuracy in pointing experiments [17]. Participants were allowed to take a break without a time limit between each block and after completing each *Task*.

5.3.5 *Data Collection*

All participant movement data was logged as a sequence of time-stamped, three-dimensional coordinates. The system also logged selection actions, whether the selection was an error and subsequent task manipulations.

Recall that the goal of this experiment is to measure the effect of different *Task*s (intended uses of the *Primary Target*) on the time taken to select the *Primary Target*. The dependent variables collected and logged by the system were selection time (T_{sel}) and errors. For all trials, we use button press down events on the *Start Target* and *Primary Target* to determine the beginning and end of the *Primary Target*'s selection movement. The selection time was the time interval from the button press action of the *Start Target* to the button press action on the *Primary Target*. The length of time to complete the *Task* was logged, but is immaterial to this experiment as we are only interested in the impact of *Task* on *Primary Target* selection time.

As noted above, the system also identified errors. We consider errors that occurred only while acquiring the *Primary Target*. Errors were classified as one of three error types:

- enter error — when participants pressed the button outside the *Primary Target* before having entered it.
- exit error — when participants pressed the button outside the *Primary Target* after having entered it at least once

- pass error — when the countdown timer reached 0 before the *Primary Target* was correctly selected.

The Error rate, displayed to participants and in our analysis, refers to the percentage of erroneous trials in a block over the total number of trials in the block (including any repeated trials). For example, if a participant failed once on every trial within a block and then succeeded on the second attempt, the error rate would be 50%.

5.3.6 *Design Summary*

We adopted a repeated-measures within-subjects design. We effectively looked at three independent variables (IVs): *Task* (TARGETING, DUALTARGETING, DOCKING, THROWING and REORIENTING), *ID* (1.49, 1.81, 2.00, 2.07, 2.37, and 2.66 bits), and *Block* (1-4). The order of *Task* was counterbalanced across participants using a Latin square [246]. Note that for each *ID*, *Primary Direction* **D** and *Secondary Direction* **D**′ were randomly ordered for generalization. The combination between **D** and *ID* and that of **D** and **D**′ were not controlled in our experiment. In summary, each participant completed 4 *Block*s × 6 *ID*s × 6 *Primary Direction*s × 5 (counterbalanced) *Task*s, i.e. 720 trials. This resulted in a data set containing 10 800 successful trials for our 15 participants.

5.4 RESULTS

We analyze our results in terms of error rate, *Primary Target* selection time (T_{sel}), motion kinematics and Fitts' Law modelling. Note that our focus is on T_{sel}, the time taken to move from the *Start Target* to the *Primary Target* given that our interest is on how *Task*s performed on the *Primary Target* impact the time to select it, T_{sel}.

5.4.1 *Error Rate*

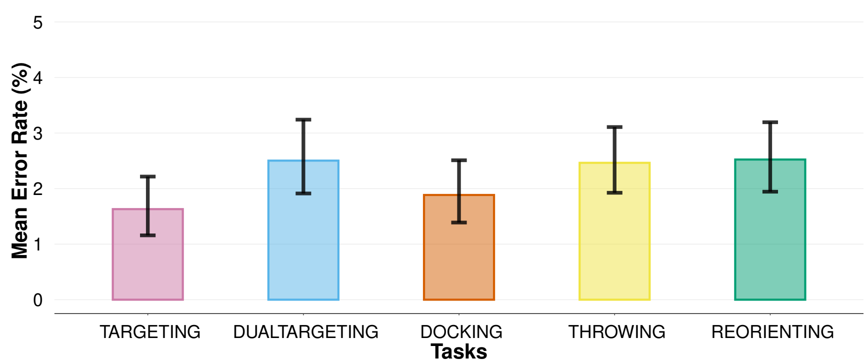


Figure 5.6: Mean error rate of T_{sel} for *Task*. Error bars are shown with 95% confidence intervals.

Recall that, in order to complete a block, participants were required to successfully complete each trial. In the case of an error in selecting the *Primary Target*, participants would need to repeat the corresponding trial at the end of the Block. As a result, alongside the 10,800 correct trials, we collected an additional 289 erroneous trials, for a total of 11,089 trials. Among 289 trials, only 3 trials had pass error and participants did not report the time pressure during the study, implying that the timer did not push participants.

Given the non-normal distribution of error rate, a Friedman test was conducted for three independent variables (IVs): *Task*, *Block*, and *ID*. We found a significant effect of *Block* on error rate ($\chi^2_{Block}(3)$=15.46, p<0.005). However, pairwise Wilcoxon rank sum tests with Bonferroni corrections did not show significant differences between blocks. With all blocks, error rate (M=2.20%, SD=5.77) was below the 5% error rate threshold that we recommended our participants not to exceed. Error rate of each *Task* was shown in Figure 5.6. The Friedman test did not reveal a significant effect of *Task* on error rate, but showed a significant effect of *ID* ($\chi^2_{ID}(3)$=12.92, p<0.05). However, pairwise comparisons did not reveal any significant differences across *ID*s.

5.4.2 *Selection Time*

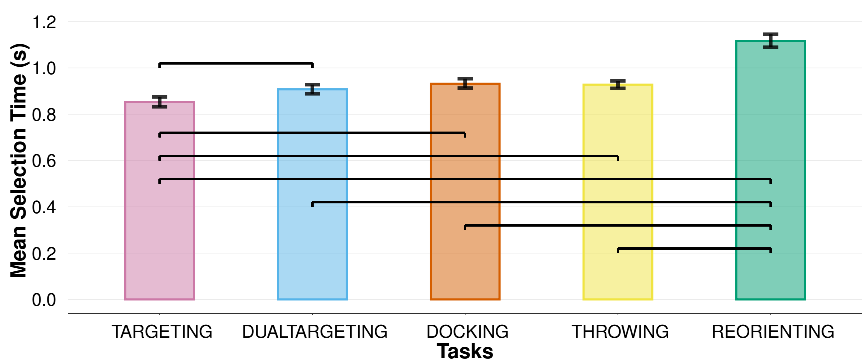


Figure 5.7: Mean selection time of T_{sel} for *Task*. Error bars are shown with 95% confidence intervals. The statistic significances evaluated by pairwise t-test are connected with lines (p<0.01).

We aggregated non-erroneous trials and removed outliers by eliminating any trial whose selection time was more than three standard deviations from the mean, leaving 10,645 trials for analysis.

Given the non-normal distribution of the data, a Box-Cox transformation [37] was applied to selection time. When sphericity was violated using Mauchly's test, Greenhouse-Geisser correction to the DoFs was applied. When significant effects were found, pairwise t-tests with Bonferroni corrections were conducted for post-hoc analysis. Effect sizes were reported as partial eta squared (η^2_p).

We first conducted a two-way RM-ANOVA (α=0.05) for selection time on *Block* and *Task* to test for a possible learning effect. We found a significant effect of *Block* ($F_{3,42}$=41.64, p<0.001, η_p^2=0.75). Pairwise comparisons revealed significant differences between Block 1 (*M*=1.03*s*) and the other three blocks (p<0.001, 2: 0.97*s*, 3: 0.94*s* & 4: 0.93*s*). We found a significant effect of *Task* ($F_{4,56}$=33.11, p<0.001, η_p^2=0.70) but we did not find a significant interaction effect between *Block* and *Task*. These results suggested a potential learning effect that did not differ between tasks; we thus removed the first block in our remaining analysis.

After removal of the first *Block*, we found a significant effect of *Task* ($F_{4,56}$=29.11, p<0.001, η_p^2=0.68) on T_{sel}. As shown in Figure 5.7, pairwise comparisons showed that TARGETING (0.85*s*) was significantly faster than the other four tasks: p<0.001 for all, DUALTARGETING (0.91*s*, 7.1% faster), DOCKING (0.93*s*, 9.4% faster), THROWING (0.93*s*, 9.4% faster) and REORIENTING (1.12*s*, 31.8% faster). Moreover, REORIENTING was found significantly slower than other tasks: p<0.001 for all. We found a significant effect of *ID* ($F_{2.1,29.7}$=254.54, p<0.001, η_p^2=0.95). Pairwise comparisons showed significant differences (p<0.01) between each ID except for 1.81&2 bits and 2.07&2.37 bits. We also found a significant interaction effect between *Task* and *ID* ($F_{20,280}$=3.00, p<0.001, η_p^2=0.18). REORIENTING was significantly slower than other tasks for all IDs (p<0.05), except THROWING at 2.66 bits (p=0.08). TARGETING was only significantly faster than DUALTARGETING and REORIENTING at 1.49 bits, and THROWING at 2.37 bits (p<0.01), as shown in Figure 5.8. There were no significant differences among other conditions.

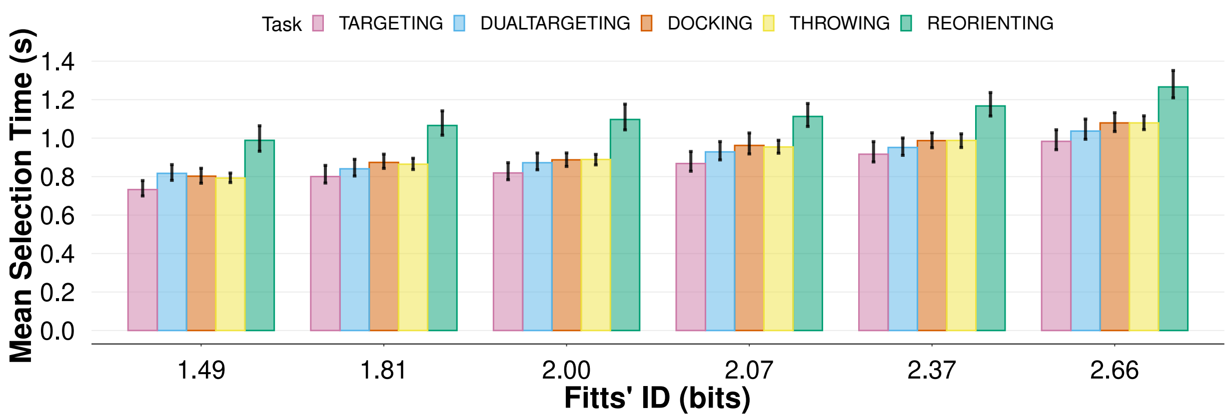


Figure 5.8: Mean selection time of T_{sel} by *ID* and *Task*. Error bars are shown with 95% confidence intervals.

5.4.3 *Motion Kinematics*

In order to try to better understand where the difference for T_{sel} between *Tasks* comes from, we analyzed the motion kinematics profile of participants' correct target selections. We once again kept all non-erroneous trials and removed the first block due to the aforementioned possible learning effect.

We computed velocity profiles based on the cursor's 3D position and corresponding timestamps. A time interval of each trial was normalized. Then, the corresponding velocities and normalized distances from the cursor

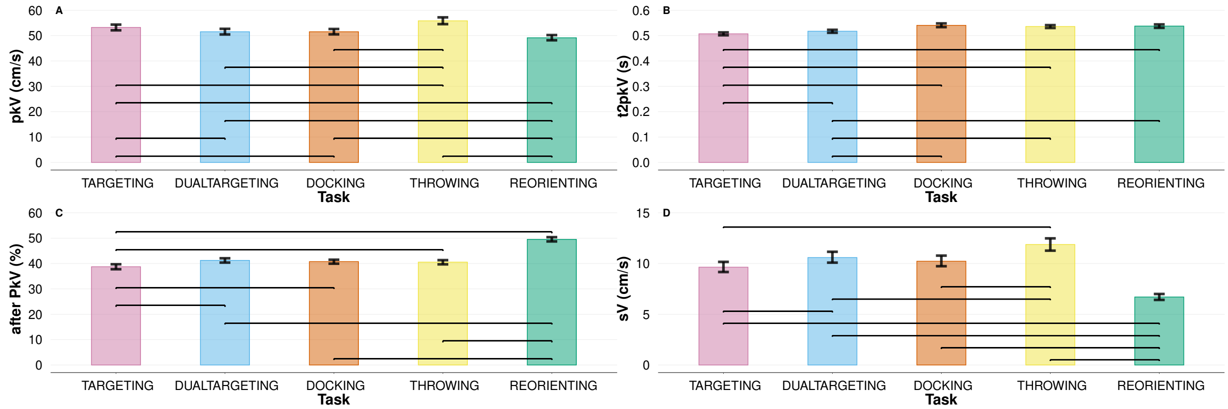


Figure 5.9: Mean pkV (A), t2pkV (B), afterpkV% (C) and sV (D) of T_{sel} for *Task*. Error bars are shown with 95% confidence intervals. The statistic significance evaluated by ART are connected with lines (p<0.01).

to the *Primary Target* were interpolated to create a time-equidistant profile (every 2% of trial time). Next, for each task and participant, these values were aggregated by computing the average for each normalized time interval. These were subsequently averaged over all participants to produce a single normalized profile for each task (see Figure 5.10).

In order to compare our results to Mandryk & Lough's work [149], we also used these profiles to compute the following motion kinematics measures: peak velocity (pkV), time to peak velocity (t2pkV), percent of time after peak velocity (afterpkV%), and selection velocity (sV: the velocity at the end of the T_{sel}) at *Primary Target* for each trial (see Figure 5.9).

For each metric, we removed outliers if the observed value was more than three standard deviations from the mean. Given the non-normal distribution of dependent variables, we conducted an Aligned Rank Transform [253] for pkV, t2pkV, afterpkV% and sV on two IVs: *Task* and *ID*. Contrasts ART [72] with Bonferroni corrections was applied as the post-hoc analysis if a significant effect was found and effect sizes were reported as partial eta squared (η_p^2).

Peak Velocity (Figure 5.9 (A)): We found a significant effect of *Task* on pkV ($F_{4,1306}$=76.67, p<0.001, η_p^2=0.19). Contrasts ART showed significant differences of pkV (p<0.001 for all) between each *Task* except between DualTargeting (M=51.52 cm/s) and Docking (51.52 cm/s). Reorienting (49.1 cm/s) was 8.3% slower than Targeting (53.2 cm/s), 4.8% slower than DualTargeting and Docking respectively, and 13.6% slower than Throwing (55.8 cm/s). We found a significant effect of *ID* on pkV ($F_{5,1306}$=608.60, p<0.001, η_p^2=0.70). Pairwise comparisons revealed significant differences (p< 0.001) between pairs of *ID* except for 1.49&2.00, 1.81&2.37 and 2.07&2.66 bits. We also found an interaction effect between *Task* and *ID* ($F_{20,1306}$=2.14, p<0.005, η_p^2=0.03). Contrasts ART also showed that there were no significant differences between 1.49&2.00, 1.81&2.37 and 2.07&2.66 bits for each task. In other words, no significant differences on pkV were revealed when target amplitudes were the same.

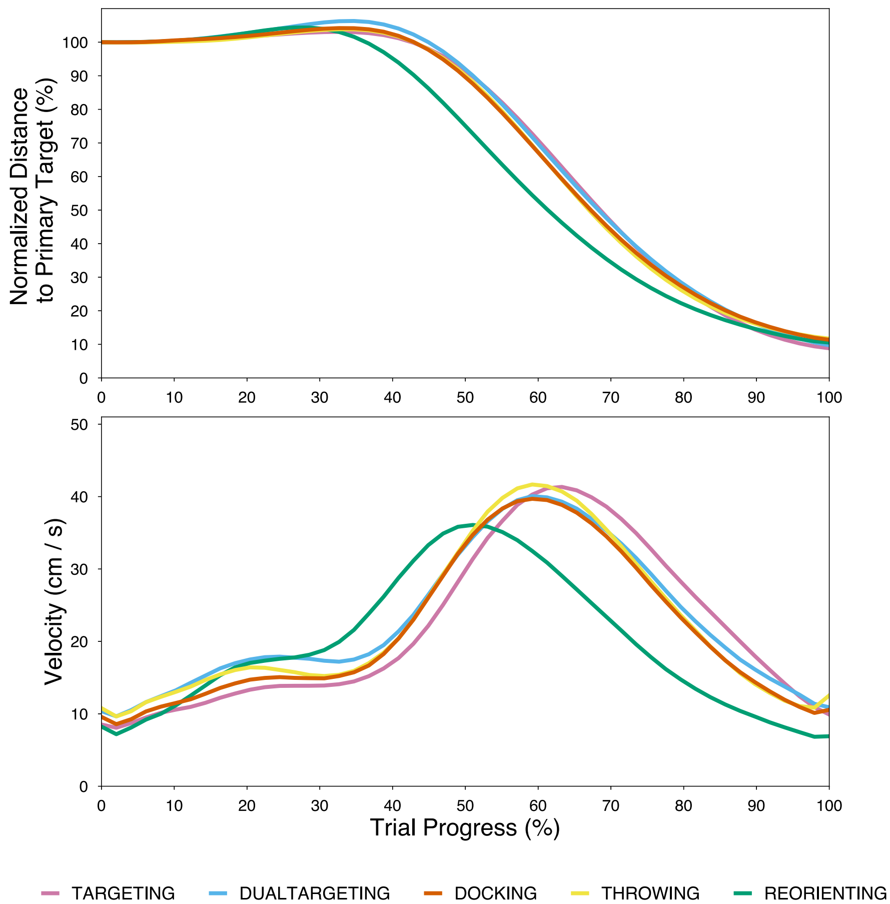


Figure 5.10: Velocity profile and normalized distance percentage to *Primary Target* in T_{sel}. Trial progress refers to the normalized time interval of a trial.

Time to Peak Velocity (Figure 5.9 (B)): We found a significant effect of *Task* on t2pkV ($F_{4,1306}$=43.31, p<0.001, η_p^2=0.12). Pairwise comparisons showed that TARGETING (0.51 s) and DUALTARGETING (0.52 s) had significantly shorter t2pkV (p<0.001 for all) than the other three tasks. More specifically, they reached the peak velocity at least 3.8% earlier than DOCKING, THROWING and REORIENTING (0.54 s for all). We found a significant effect of *ID* ($F_{4,1306}$=27.56, p<0.001, η_p^2=0.10). Pairwise comparisons showed significant differences (p< 0.001) between each *ID* except for 1.49&2.00, 1.81&2.37, 2.07&2.37 and 2.37&2.66 bits. We did not find an interaction effect between *Task* and *ID*.

Percent of Time after Peak Velocity (Figure 5.9 (C)): A significant effect of *Task* was revealed on afterpkV% ($F_{4,1306}$=182.49, p<0.001, η_p^2=0.36). The post-hoc analysis showed TARGETING (38.7%) had significantly smaller afterpkV% (p<0.001 for all) than the other four tasks while REORIENTING (49.5%) had significantly larger afterpkV% (p<0.001 for all). Participants spent 28% more time after peak velocity in REORIENTING compared to TARGETING and at least 20% longer than other three tasks, suggesting a longer deceleration phrase in REORIENTING. We found a significant effect of *ID* on afterpkV% ($F_{4,1306}$=200.95, p<0.001, η_p^2=0.43). Pairwise comparisons showed significant differences (p< 0.001) between each *ID* except for 2.00&2.07 bits. We did not find an interaction effect between *Task* and *ID*.

Selection Velocity (Figure 5.9 (D)): We found a significant effect of *Task* on sV ($F_{4,1306}$=123.43, p<0.001, η_p^2=0.27) and the contrasts ART revealed that REORIENTING had significantly slower sV (7.70 cm/s, p<0.001 for all) and THROWING had significantly higher sV (11.87 cm/s, p<0.001 for all) than other tasks. Compared to other tasks, sV for REORIENTING was at least 25% slower and over 50% slower than THROWING. DUALTARGETING (10.59 cm/s) also had significantly higher sV than TARGETING (9.64 cm/s, p<0.01, 9.9% higher). We found a significant effect of *ID* on sV ($F_{5,1306}$=95.62, p<0.001, η_p^2=0.27). Pairwise comparisons showed significant differences (p< 0.01) between each *ID* except for 1.49&1.81, 1.49&2.07, 1.81&2.07, 2.00&2.37 and 2.37&2.66 bits. We also found an interaction effect between *Task* and *ID* ($F_{20,1306}$=2.51, p<0.001, η_p^2=0.04). Pairwise comparisons showed that there were no significant differences between 1.49&1.81, 1.49&2.07, 1.81&2.07, 2.00&2.37, 2.37&2.66 and 2.00&2.66 bits in each task. Interpreting these numbers in terms of target width and amplitude, these results argue that sV is significantly impacted by target width.

5.4.4 *Fitts' Law*

Given our research question, i.e. whether *Task* users will perform on an object impacts the time taken to acquire the object, Fitts' law modelling and throughput analysis were applied only to the selection of *Primary Target*, i.e., only to T_{sel} in Figure 5.2 but not to the subsequent manipulation task. To perform our analysis, we computed the effective target width (W_e) for each target width by multiplying the standard deviation by 4.133 [146, 165] and used W_e to calculate the effective ID, ID_e accordingly. As a result, the difference between W

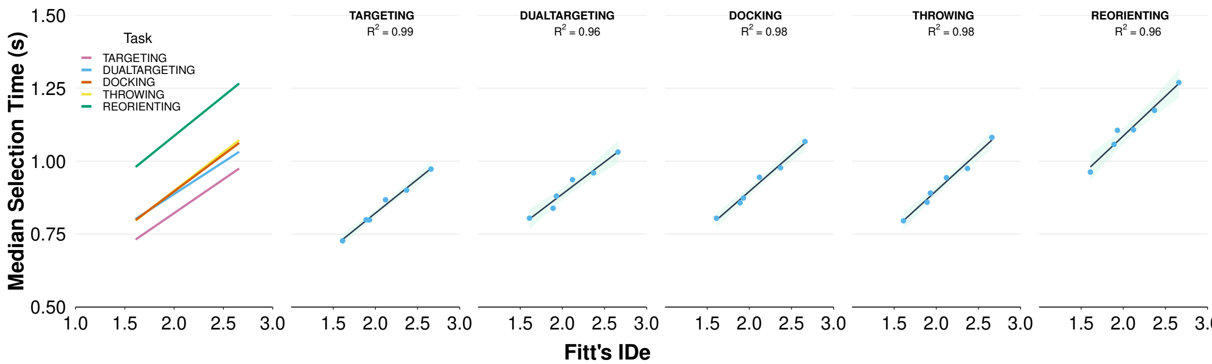


Figure 5.11: Median Selection Time as a function of Fitts' ID_e per task, with corresponding R^2 and 95% confidence interval (light-green area).

and W_e ranges from 0 cm to 0.6 cm and the corresponding difference between ID and ID_e ranges from 0 bits to 0.12 bits, as shown in Table 5.1.

The classical Fitts' Law was used for modelling [82, 145] due to two concerns: 1. for ID, as our targets are spheres, Fitts' law variants for targets in arbitrary shapes are not necessary. 2. as Triantafyllidis & Li [227] points out, no work has included all spatial factors in 3D space and a standard metric for 3D modelling is missing. Meanwhile, the classical formula is still a common practice in VR [21, 225, 227].

A(cm)	W(cm)	W_e(cm)	ID(bits)	ID_e(bits)
9.0	3.0	3.2	2.00	1.93
9.0	5.0	4.4	1.49	1.61
12.5	3.0	3.0	2.37	2.37
12.5	5.0	4.6	1.81	1.89
16.0	3.0	3.0	2.66	2.66
16.0	5.0	4.8	2.07	2.12

Table 5.1: Effective target width (W_e) and effective ID (ID_e) are calculated for each pair of target amplitude (A) and target width (W).

To compensate for the non-normal distribution of selection time in T_{sel}, we adopted a common practice [57, 214] where we first aggregated the mean for each effective Index of Difficulty (ID_e) per *Block* per *Task* and then aggregated the median for each ID_e and *Task*. ID_e ranged from 1.61 to 2.66 bits and the aggregated median time of all tasks correlate with ID_e positively ($R^2 \geq 0.96$), as shown in Figure 5.11. When looking at the coefficients of these linear regression models in Table 5.2, we noticed that slope values b were relatively similar across *Task* but Reorienting has a higher y-intercept value a than the other four tasks. We also report throughput scores in Table 5.2. Unsurprisingly, Targeting had the largest throughput, Reorienting had the lowest throughput, and the throughputs of the other three tasks were similar, a result consistent with T_{sel} values.

Task	a	b	TP
TARGETING	0.36 [0.28 0.43]	0.23 [0.20 0.28]	2.47 [2.34 2.61]
DUALTARGETING	0.44 [0.32 0.57]	0.22 [0.16 0.28]	2.29 [2.15 2.44]
DOCKING	0.39 [0.30 0.49]	0.25 [0.21 0.30]	2.25 [2.13 2.38]
THROWING	0.37 [0.27 0.48]	0.26 [0.21 0.31]	2.26 [2.13 2.38]
REORIENTING	0.54 [0.38 0.70]	0.27 [0.20 0.35]	1.87 [1.76 1.99]

Table 5.2: Modelling results of Fitts' Law ($MT = a + b \cdot ID_e$) and throughput values $TP = ID_e / MT$: estimates (95% CI)

5.5 GENERAL DISCUSSION

The goal of this paper is to explore the impact of intended use of a target, i.e. the *Task* performed on the target, on the time required to select that target, i.e. T_{sel}, before performing the manipulation. We examined five tasks: classical TARGETING, DUALTARGETING, DOCKING, THROWING, and REORIENTING. Our hypothesis is that there is an impact, i.e. that the selection that precedes target manipulation is impacted by the specific manipulation *Task* that we perform on the selected target.

Our results support this hypothesis. While we did not find a significant effect on error rate, we did find significant differences in the time taken to select the target. In particular, we found that selection preceding classical TARGETING took significantly less time than selections preceding all other *Task*s we tested, and that selection preceding a REORIENTING *Task* took significantly longer than selections preceding all other tasks. More specifically, with the interaction effect between *Task* and *ID*, TARGETING was significantly faster than some *Task*s at certain *ID*s and REORIENTING was significantly slower than all other *Task*s at all *ID*s except for THROWING at 2.66 bits. Target selection preceding DUALTARGETING, DOCKING and THROWING did not differ significantly in the time taken. While differences of selection time between those four tasks and TARGETING were small (within $0.27s$), they had been already more than 7% slower, with REORIENTING over 30% slower than TARGETING. Given that there are significant differences for TARGETING and REORIENTING, we reject our null hypothesis (no impact of intended use) and claim evidence for the impact of intended use on target selection.

To understand where and how this difference occurred, we analyzed movement time during the selection task. First, we created kinematic profiles of distance and speed during target selection. A visual inspection of Figure 5.10 shows that REORIENTING has a lower peak speed and reaches peak speed earlier in movement with a corresponding increase in the length of time spent decelerating. Further kinematic analysis supports this visual analysis; in Figure 5.9, REORIENTING has the lowest peak speed (at least 4.8% lower) and the longest deceleration phase (more than 20% greater) than other *Task*s by a significant margin. It implied that participants planned their movement and did not rush during the selection for REORIENTING. Besides, the Fitts' law modelling confirmed this by showing that the reaction time for REORIENTING was higher and the throughput value of REORIENTING was smaller than other tasks.

The kinematic analysis does, however, present some additional observations that merit future investigation. For example, REORIENTING results in the lowest peak speed; THROWING results in the highest peak speed in the preceding selection movement, which is followed by peak speed for selection preceding TARGETING. DUALTARGETING and DOCKING do not result in significant differences in peak speed in the preceding selection kinematics. Effect sizes are not small for various measures of kinematics highlighted in Section 5.4.3. Based on typical interpretations of effect size, $\eta^2 > 0.14$, i.e. large effect size[1], for all differences except time to peak speed ($t2pkV$), where the effect size maps to medium effect. The absence of additional significant differences in selection time between DUALTARGETING, DOCKING and THROWING does not imply that differences in selection prior to these *Task*s do not exist; it simply implies that we measured no significant differences in selection time. Future work is planned to probe these effects in more depth.

It is interesting to contrast our results in Mandryk's and Lough's results in 2D [149]. Mandryk's and Lough's four intended uses (*Task*s) correspond to TARGETING, DUALTARGETING, DOCKING, and THROWING in our experiment. They found that TARGETING and DUALTARGETING resulted in selection times (T_{sel}) that differed significantly from DOCKING's and THROWING's selection times, but did not observe differences between TARGETING and DUALTARGETING. In contrast, while we, too, found that TARGETING resulted in the shortest preceding selection times, DUALTARGETING resulted in T_{sel} more similar to DOCKING's and THROWING's T_{sel}. In terms of overall differences, we and Mandryk and Lough find differences in T_{sel} under 10% in overall magnitude for these selections (9.4% in ours and 8.8% estimated in theirs). Our REORIENTING is unique to our study, and resulted in the most significant differences in preceding selection time; REORIENTING resulted in target selection times more than 30% longer than TARGETING.

Given our results that the intended use of a target impacts the time taken to select a target, the question that follows from this is what these results

1 https://www.spss-tutorials.com/effect-size/

mean. To address this question, we point, again, to the analysis of 20 years of VR-based by Bergström et al. [25]. As noted earlier in our review of this work, Bergström et al. highlight that, as dependent variables, selection task studies measure selection time for the selection task and manipulation task studies measure the time participants take to manipulate an object. Success in selection-based research is measured by shortening selection times or reducing errors, or both; success in manipulation is similarly based on increased throughput for the manipulation task. The assumption that underlies these success metrics is, *ipso facto*, that each individual user action can be optimized in isolation from other tasks in interfaces, i.e., that manipulation does not impinge upon preceding selection, but our results argue that this assumption cannot be made. If different manipulations impact the time taken for a preceding selection, then measuring only the manipulation time may over-estimate (or under-estimate) the benefits of a novel interaction technique.

It is also true that we only measure retrospective impact (i.e. the impact of future intended use on preceding selection), but prospective impacts are also possible. It is hypothetically possible that a pointing facilitation technique in VR, e.g. an area cursor, might impact a user's ability to perform a task on a target, e.g. reorientation of the target acquired via the area cursor. While exploring prospective impacts is one area of future work, it also highlights a more general implication for system design. Specifically, when we have new interaction techniques (i.e. new manipulations) or new pointing facilitation techniques (i.e. new selection techniques), incorporating them into realistic systems requires thinking not just about the individual action that they optimize but also about their place within and more general impacts on the overall task flow of the user.

5.5.1 *Applicable Scenarios*

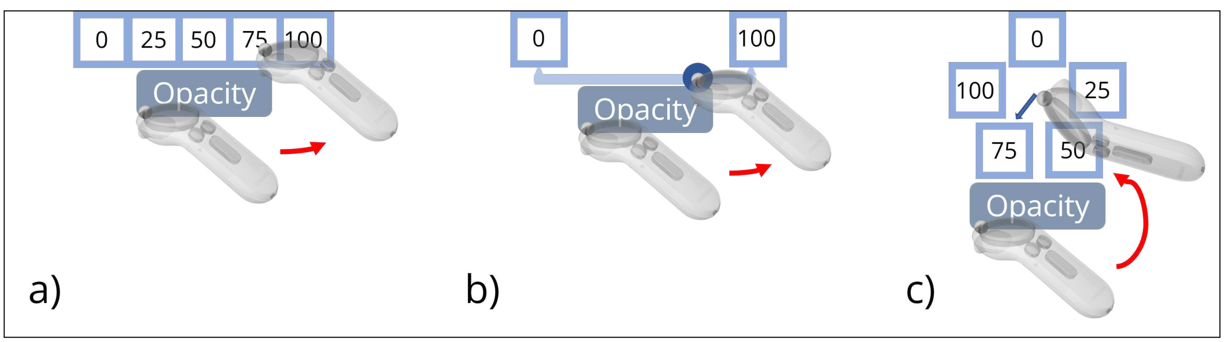


Figure 5.12: Menu selection for changing target opacity: a) DUAL-TARGETING, b) DOCKING, c) REORIENTING.

Our results can be framed into concrete applications and interface design in VR. One classical example is menu selection [80, 164]. Considering Figure 5.12, our results suggest that interactions that take advantage of DUALTARGETING and DOCKING may result in similar selection time prior to the manipulation sub-task and differ from each other based on the design of manipulation techniques. In contrast, techniques that leverage REORIENTING already take longer selection time prior to the manipulation. Similarly, in data visualization tasks in virtual environments [40], rotational manipulations of a dataset may introduce additional costs if the target acquisition prior to the rotational manipulation is slowed.

5.5.2 *Limitations*

One highlighted limitation is that the range of IDs (1.49 to 2.66 bits) in this experiment is fairly small for a Fitts' Law design. This is because we restrict our current experiment to arms' length interactions with a controller on static targets, so target widths and distances are constrained for a reachability concern in VR. While these IDs are commonly used in an arms' length interactions in VR environment, they are low compared to desktop interfaces [149], touch-based interfaces [90], distant interactions in VR [17, 21, 57, 58], contexts where movement amplitudes can increase due to the greater distance of targets from the user. This explains why the throughput values in Table 5.2 are relatively lower, compared to throughput scores in other VR studies [21, 225]. However, incorporating more distant interactions adds additional complexity to the selection action because direction (e.g. targeting via a ray) and depth are often controlled differently during distant interaction [17, 21]. Future work could assess the findings for a wider range of IDs.

While we contrast five different *Task*s (the intended use) in our study, there exist more complex manipulations in the virtual environments. Furthermore, we do not consider objects' surface characteristics, perceived weight, and perceived fragility [113, 198] as objects in our study have similar surface, weight and fragility. We also do no consider bimanual interactions [95] nor co-articulated actions (e.g. 6-dof reorienting, i.e. a docking task that requires both rotation and translation) [112, 201]. As noted in the experiment design, we did not control the *Primary Direction* and *Secondary Direction*. Machuca and Stuerzlinger [20] found that target acquisition in virtual environments was slower and had less throughput along the depth-axis, i.e. Forward and Backward in our experiment, than lateral directions. Further research can explore how directions can impact the acquisition for various intended uses, particularly for co-articulated actions.

Finally, it is noted that our experiment was conducted remotely in participants' homes. While an in-home environment increases the external validity of our study, it cannot be as controlled as a laboratory one, whose setup and control can assume to be optimized for the experiment. Space in homes may be constrained, and households may present interruptions during the experiment. To limit this as a factor, we note that participants were provided with detailed instructions, from both pre-recorded videos and experiment systems. A researcher was also present via video conferencing tools during the experiment, allowing the environment to be monitored for confounds. We also note that our within-subjects design partially controls for confounds by ensuring that the environment is similar across each task for a participant.

5.5.3 *Future Work*

The results of our intended use study present interesting avenues of future research into interaction in virtual reality environments. As one example, a novel interaction technique might not result in higher throughput during the interaction, but it is possible that it might speed the selection that precedes the interaction. In this case, considering both the selection and manipulation of a target as a unified task allows us to identify potentially beneficial novel interactions that might have been ignored if the only metric for success is throughput for the manipulation.

Another possible area of inquiry given differences in kinematics noted in Section 5.4.3 is that a more careful analysis of movement during selection might allow the system to infer what a user intends to do with a target prior to acquiring the target. This, in turn, could allow us to develop interaction techniques that leverage this inference. Reorientation appears a good initial candidate to identify given the deviation in the selection kinematics shown in Section 5.4.3.

5.6 CONCLUSION

In this chapter, we investigate how what users intend to do with a virtual target impacts their initial acquisition behavior (RQ5). We question an assumption in past research that the acquisition and the subsequent manipulation can be independently optimized [25]. Based upon past work in 2D environments, we examine the impact of five common virtual reality manipulations (TARGETING, DUALTARGETING, DOCKING, THROWING, and REORIENTING) using the time taken to select a target prior to performing the manipulation. We identify the existence of an effect of intended use of a target on the act of selecting the target. Specifically, we find that TARGETING had the shortest selection time, REORIENTING had the longest, and the other three intended uses we evaluate result in acquisition times between these two extremes. We synthesize these results and highlight their implications for research and design in VR environments, which can possibly be applied to real 3D manipulation tasks[231], such as the direct manipulation of the lamp.

6 CONCLUSION

In this chapter, after summarizing the contributions of this thesis, I discuss the design challenges and technical limitations of our prototype and controlled experiments. The potential impacts of each work are highlighted and future research projects are proposed for more pervasive desktop computing.

6.1 SUMMARY OF CONTRIBUTIONS

I summarize my contributions using the taxonomy on research contributions in Human-Computer Interaction proposed by Wobbrock and Kientz [254].

6.1.1 *Artifact Contributions*

In Chapter 3, we developed a novel direct manipulation SAR system (LuxAR) in the form of an augmented architect desk lamp. This prototype integrates a pico laser projector, position and orientation tracking, and allows for direct manipulation to extend and augment conventional desktop computing. LuxAR enables bidirectional connection to transfer content between devices and the surrounding environments, and adapts content representation based on surfaces, objects, and other devices. Meanwhile, it supports multiple interaction methods that are commonly found in the existing desktop computing environment.

6.1.2 *Empirical Contributions*

In Chapter 3, we conducted a user study with the LuxAR prototype, including semi-structured interviews to examine proposed interactions and consider potential scenarios and applications. This empirical work led to design considerations for direct manipulation systems in augmented desktop computing.

In Chapter 4, we conducted a controlled experiment in Virtual Reality to investigate coupled and decoupled target acquisition methods in dynamic peephole interaction on an augmented desk. This study provided insights into the performance, user preferences, and strategies for different interaction methods in augmented desktop environments.

In Chapter 5, we conducted a study examining the effects of intended use on target acquisition in virtual reality environments, considering five different manipulation tasks: targeting, dual-targeting, throwing, docking, and reorienting. This empirical work demonstrated how the intended use of a target affects its acquisition time and movement patterns in VR environments, shedding insights on lamp acquisition and its subsequent direct manipulation.

6.2 APPROACHING RESEARCH QUESTIONS RETROSPECTIVELY

The goal of this thesis was to make desktop computing more pervasive. To achieve this goal, we answered research questions in progressive steps.

Our first research question (RQ1) focuses on creating bidirectional connection across displays, devices, and physical surfaces, enabling content to move seamlessly between screens and environments. We propose LuxAR, a Spatial Augmented Reality (SAR) system using an architect lamp, designing interaction techniques based on its physical properties. These techniques enable users to interact with digital information and enhance physical spaces using both the augmented lamp as an input and other inputs commonly found in the workspace, such as mouse, pen, direct touch, and other devices (RQ2). Through the direct manipulation of the augmented lamp, users can transfer existing applications, like a virtual calendar, from the desktop screen to surfaces such as a desk, wall, or ceiling, and back again. When an application is on a surface, its content adapts accordingly. Users can then adjust the lamp's posture to interact with the content or use other input methods to engage with it. These interactions greatly enhance our current desktop computing experiences (RQ3). These interactive experiences motivate new questions, where we look at the dynamic peephole interaction created by the direct manipulation of our augmented lamp on surfaces. Users need to use at least one hand to manipulate the lamp, leading us to investigate how coupled and decoupled control mechanisms impact target acquisition in dynamic peephole interaction on the augmented surfaces (RQ4). While these two control mechanisms reveal marginal differences in performance metrics, they exhibit distinct characteristics in peephole velocity and target revelation within the same search task. Additionally, participants adopt different interaction strategies in the dynamic peephole acquisition task. Based on this, we explore the broader question of how a user's intentions toward virtual targets influence the initial acquisition of these targets (RQ5).

We now take a retrospective step to reflect and examine how later findings can help to improve and refine previous designs and questions. Our findings in Chapter 5 reveal that the intended interactions of users with a virtual target affect its initial acquisition and related motion kinematics, including the final acquisition velocity. We conclude that individual actions cannot be optimized in isolation from other tasks. In Chapter 4, we find marginal differences in performance metrics between coupled and decoupled methods. As mentioned in the discussion of Chapter 4, participants continuously held the virtual lamp during the pointing task and did not release the lamp until the break. In contrast, the setup described in Chapter 3 involved time gaps between tasks, which required participants to grab and release the physical lamp repeatedly. This continuous grabbing interaction was necessary to reduce the duration of the experiment, but it is not representative of the real use we envision, consisting of intermittent interactions with the lamp. If we take these into account, the overall interaction time for the coupled and decoupled control mechanisms for an acquisition task in dynamic peephole may differ. This may need further efforts to refine our design guidelines for coupled and decoupled control with a physical lamp.

Our findings described in Chapter 5 are obtained in a virtual environment using controllers, but they have the potentials to be applied to our design of LuxAR in Chapter 3. Key findings applicable to our LuxAR prototype are that peak velocities and final acquisition velocities vary according to the intended use. This implies that when users intend to rotate the lamp, adjust the lamp's height, or simply grab it, their approaching and grabbing velocities may differ. A recent work by Valkov et al. has leveraged these motion characteristics of fingers to predict the object a user intends to grasp or is currently holding [231]. Therefore, we can potentially use these motion kinematic features to predict a user's intentions. To obtain these motion features, we can integrate different sensors, including force sensors [81, 175], inertial measurement units (IMUs) [213], ultra-wideband sensors [74], and proximity sensors [77], into the lamp shade or wearable devices like smartwatches. This integration would enhance the lamp's awareness of user activities, allowing it to provide appropriate interaction techniques and physical manipulation capabilities. For instance, an enhanced lamp could selectively enable or disable specific manipulation axes based on user intent, such as adjusting lamp height to change the view of a calendar application or rotating the lamp to move it to other surfaces. This adaptive design provides stable support along certain axes while allowing more flexible manipulation along others, tailoring the interaction to the user's needs and the task at hand.

Considering our results in Chapter 4, there was no one-condition-fits-all approach for the acquisition of the target in the dynamic peephole. We found subtle differences in accuracy, total time, and search time between coupled and decoupled control mechanisms, with coupled method showing significantly faster acquisition time. The coupled method is preferred for its convenience, and reduced physical load. On the other hand, the decoupled method with direct touch is preferred for its accuracy. If our prototype was free from current mechanical constraints and as manipulable as the virtual lamp, we could seamlessly switch between acquisition techniques without a concern for overhead costs. This flexibility would be particularly beneficial in scenarios described in Chapter 3. For high-precision tasks, such as interacting with small calendar buttons or augmented views on desk-placed smartphones, direct touch would be ideal as soon as the information is revealed by the lamp. For acquiring larger pieces of digital information (like application windows) or performing extended manipulations (such as moving information around), using the lamp itself would prove more comfortable and efficient. This adaptive approach aligns with our earlier findings on the importance of considering task context and user intentions in optimizing interaction techniques. It also reinforces the need to examine how continuous versus discrete controls affect performance across various tasks, as discussed in relation to our findings in Chapter 5.

6.3 pervasive desktop computing beyond an augmented lamp

We addressed our research objective by building a small, personal SAR system into a modified architect lamp. But is this augmented lamp still a "lamp"? The core purpose of a lamp is to provide illumination. In our augmented lamp, we replace the light bulb with a projector, attach markers onto the lamp shade for tracking, and integrate a button to enable direct control from the lamp. From a philosophical standpoint, we could argue that the essence of "lampness" is preserved, as the augmented lamp still fulfills a purpose of "lighting", in terms of illuminating an (interactive) surface. Meanwhile, this illumination has a more focused beam of light, and is much less luminescent due to the laser projection. Beyond that, our modifications to the architect lamp transform it into a multi-functional input and output device that extends its utility far beyond that of a conventional lamp, and bridges the gap between digital and physical environments. With a projector, it displays the digital information on physical surfaces; with a button, it enables users to pick and drop the digital information; with cameras and its unique physical characteristics, it enables bidirectional connection between *Environment* dimensions, augments the physical surfaces and objects, understand users activities, and supports long-duration tasks with diverse inputs. In the context of pervasive computing, where the goal is to integrate computing capabilities into everyday objects and environments and become aware of users activities, the augmented lamp exemplifies this blend of physical and digital worlds. It maintains its physical form and primary function while incorporating advanced features that make it an active participant in the pervasive desktop computing environment. Therefore, it is still fundamentally a lamp, but also has evolved into a hybrid device that embodies the principles of pervasive computing.

Despite the advantages highlighted previously, it is worth exploring other solutions as options to make desktop computing pervasive and enable the bidirectional connection across the *Environment*. As discussed in Chapter 2, HMDs for AR and MR have been used to extend virtual information on static surfaces [137] and enable dynamic information placement in physical environments [168]. While current HMDs are relatively heavy for extended use (*e.g.* Oculus Quest 3 at 515 grams, Apple Vision Pro over 600 grams [7]), the trend towards lighter, more minimal designs [78] presents exciting future opportunities. A unique aspect of LuxAR is the use of an architect lamp as an input device. Its mechanical properties allow for flexible manipulation while maintaining stability afterwards. This concept could be combined with HMDs to achieve similar interaction techniques as described in Chapter 3, where lamp interactions in physical space represent equivalent mouse interactions in traditional desktop environments. It should be noted that the lamp design is not the only form factor supporting such functionalities. Other systems, like the steerable chair in Joshi et al.'s work [116] or the LuminAR system with a robotic lamp [139], demonstrate similar potential for workspace augmentation. The key distinction lies in the explicit and direct manipulation of the physical objects, directly linking a user's physical activities to their intentions. This approach is closely aligned with the concept of tangible user interfaces [110], where the manipulation of physical objects can significantly improve user performance and digital information experience. Examples include engaging digital information through grasp-based [224] and rotation-based [174] inputs to physical objects, potentially lowering the cognitive load associated with complex digital tasks, and enhancing the overall user experience.

6.4 THE IMPORTANCE OF BIDIRECTIONAL CONNECTION

Recall that the bidirectional connection describes the ability of a system that moves information across these dimensions while preserving the awareness during this movement. Its importance in pervasive desktop computing lies in empowering users with choices that suit their diverse needs across different contexts. While much effort has been devoted to designing interaction techniques and systems in a one-size-fits-all way, our research reveals that user needs can vary significantly depending on the context. Our findings in Chapter 3 show that different applications and interactions showed varying levels of appreciation depending on their context and placement. For example, the music player was well-received when projected on the ceiling, while the calendar application was less effective in this position. Similarly, the lamp height adjustment interaction proved suitable for manipulating the floor plan but was indeed less practical for the calendar application. Our results in Chapter 4 also demonstrate that users prefer decoupled methods when accuracy is paramount, but opt for coupled methods when speed is the priority. This variability in user preferences highlights the need for flexible systems that can adapt to different interaction contexts. Pervasive computing emphasizes the importance of embedding computing capabilities into the physical environment and understanding user activities. Meanwhile, we still recognize that traditional desktop computing remains as a central hub for various digital tasks, including coding, interface design, writing, and information browsing. Therefore, for a pervasive desktop computing environment, users should have the options and power to decide what, where and how information is present to them. Thus, this bidirectional connection allows users to fully leverage the benefits of a pervasive desktop computing environment, and tailor their interaction methods flexibly, thereby enhancing their overall computing experience and productivity.

6.5 LIMITATIONS

While this thesis provides results to demonstrate the possibility of using an augmented lamp system to enable pervasive desktop computing, and investigates user behaviors on an augmented surface, there are several limitations, from either a technical implementation perspective or design and methodology perspective that might require further investigation.

We considered an ideal desktop workspace, which may not always be the case. Our virtual workspace was reconstructed based on carefully configured physical workspace parameters. We strategically placed and registered physical objects at fixed locations within the system, maintaining a clean desk environment with essential registered items, such as the keyboard and smartphones. To optimize visual presentation, we primarily displayed virtual content on empty or available surfaces. While our results demonstrated the system's feasibility in establishing a bidirectional connection among *Environment*s and adapt content based on the understanding of the physical environment in a controlled lab setting, we need to consider addressing more complex workspaces, such as those illustrated in Figure 3.2, dynamically and flexibly. This includes content adaptation based on understanding the geometric aspect of the physical environment [99, 194], as well as accounting for user actions and behaviors [140, 168, 259]. These considerations present opportunities to enhance the system's adaptability and robustness in diverse, real-world settings.

We primarily focused on the acquisition task to gain insights into user behaviors within augmented workspaces. However, it is important to acknowledge that both traditional desktop computing and augmented desk environments encompass a diverse range of tasks. For instance, in the user study of LuxAR, participants were asked to pick-drag-drop virtual windows at different locations (a docking task), and they were also asked to follow visual cues to find a virtual window (a navigation task). As emphasized at the beginning of this thesis and further elaborated in Chapter 4, the introduction of LuxAR has opened up novel interactive experiences and created new opportunities for interactions. As acquisition is the first step for sequential interactions, we felt it was crucial to address the questions related to this action. By investigating it, we aimed to establish a solid understanding for more complex interaction patterns that build upon this initial engagement. This focus does not diminish the importance of other tasks, but rather serves as a starting point for a more comprehensive understanding of user interactions in this new augmented workspace.

Our results in virtual world may not generalize to a physical world, but the use of a fully virtual environment in this thesis provided an ideal controlled environment for our experiments. In the previous two chapters, we ran studies in VR rather than directly in the physical world, allowing us to isolate and examine specific factors more precisely. For instance, if we had used LuxAR directly for target acquisition in dynamic peephole interaction, the lamp's mechanical limitations, such as constrained manipulation angles, and potential distortions in the virtual display, could have introduced confounding effects that would have been difficult to control and explain. Similarly, our study on the effects of intended use on target acquisition benefited from the controlled VR environment. As highlighted in that chapter, the surface characteristics of physical objects, perceived weight, and perceived fragility could have introduced additional variables, potentially confounding our results [113, 198]. By conducting these experiments in VR, we aimed to achieve better control over experimental conditions and produce results that could be more clearly interpreted and explained. We could then bridge this gap by gradually introducing physical-world complexities, allowing for a better understanding of how these results translate to real-world scenarios. This staged approach, from controlled virtual environments to more complex physical settings, can also help build a comprehensive understanding of user interactions in augmented workspace.

6.6 FUTURE DIRECTIONS

We describe future research opportunities that can be pursued. We provide initial ideas of their design, technical implementation, and possible user studies required to understand how people perceive the proposed interaction techniques.

6.6.1 *Short-term Projects*

A self-contained LuxAR with Intended Use Prediction

A self-contained LuxAR system has the potential to create a more portable and flexible solution for pervasive desktop computing. To achieve this, we propose replacing the OptiTrack system with an integrated suite of sensors: an RGB-Depth camera, tracking camera, and Inertial Measurement Units (IMUs). These could be supported by simultaneous localization and mapping (SLAM) algorithms [44] for accurate pose, depth, and trajectory estimation, as well as scene reconstruction using point clouds. To enhance environmental understanding and adaptive display of virtual information, we could implement vision-based models and algorithms for object detection [34, 191] and segmentation [103, 125], enhancing the understanding of the physical workspace and allowing for faster prototyping. Besides IMUs, a variety of sensors can also be integrated into the lamp's shade, including force sensors, ultra-wideband sensors, and proximity sensors. This sensor array would significantly improve the system's awareness of user activities by leveraging the motion kinematics of users' hands. Therefore, we can predict intended manipulations and interactions, allowing the system to proactively highlight relevant content and interfaces. With these capabilities, a self-contained system, complemented by a compatible toolkit, would empower practitioners to design and enhance personalized desktop experiences.

LuxAR with Bi-manual Interaction in Dynamic Peephole

Future research could significantly expand the interaction possibilities by exploring bi-manual interactions in dynamic peephole scenarios. This idea would involve using the lamp with one hand while employing various complementary input modalities with the other, including pen-based interactions [45, 217], direct touch [258], mouse inputs [9, 218], and hand gestures [139, 251, 259]. By investigating these multi-modal interactions across diverse physical surfaces and displays, we could substantially enrich our design spaces for a pervasive desktop computing workspace. For instance, once the lamp has acquired virtual information, we could design interaction techniques based on the presence of secondary input methods, whether it is directly on the virtual information, positioned above it, or even located on a separate surface, and conduct a controlled experiment to investigate how a dynamic peephole could potentially impact these techniques.

A longitudinal User Study

Unlike previous research projects, which focus on ad hoc interactions with the augmented lamp, this work focuses on the user experience and feelings when using the augmented lamp over time. Similar to prior projector-based [101] and tabletop-based [245] projects, this project studies what kind of tasks and information users prefer interacting with, where (*i.e.* which surfaces) they want to anchor and engage with the information, when they want to interact with it, and how long they want the information to be alive and available in the long term. A diary method with system logs is preferred. We intend to investigate the benefits and drawbacks of using the augmented lamp, with a particular emphasis on the long-term effects. We also hope to learn where users spatially distribute virtual information in the environment, which applications are favored and will be used for an extended period of time, and how users will adjust their routines and behaviors to socially accept the augmented lamp.

6.6.2 *Long-Term Projects*

Adaptive Content Placement and Interaction for Occlusion

As highlighted in Section 6.5, our prototype takes mainly advantage of all available flat surfaces. The desktop workspace is always an occluded and complex environment where the occlusion for virtual information not only comes from physical objects, but also from users [86, 235]. Therefore, when we transfer content from one surface or device to another surface or device, the presence of objects can partially occlude the virtual content (*e.g.* content distorted on a cup) and nonplanar and irregular shapes of surfaces distort the content (*e.g.* edges between desks and walls); when users interact with such information, the information would be overlaid on hands. The content should adapt to it and maintain visual aesthetics and usability when meeting these scenarios.

An exploratory study revealed that users do not mind irregular projected displays for knowledge work with conventional interfaces [237]. Therefore, it is important to allow content to naturally address these irregular occlusions. The idea of "Freeform Interfaces" was proposed by Serrano, Roudaut, and Irani [210] and many works have been proposed to reconfigure the interfaces layout based on the geometric shapes of physical objects [99] or surfaces [168, 169, 194]. However, as both the projected display and content can be dynamically manipulated in space, this presents a different context for interactions, and it remains unclear how the content layouts should dynamically change to address those concerns. In addition, when users interact with the already adapted content, it is interesting to see how this content should be further adapted to avoid the occlusion introduced by the input, *e.g.* hands, and how users would further organize both their workspace and digital content for a better cohabitation environment [259].

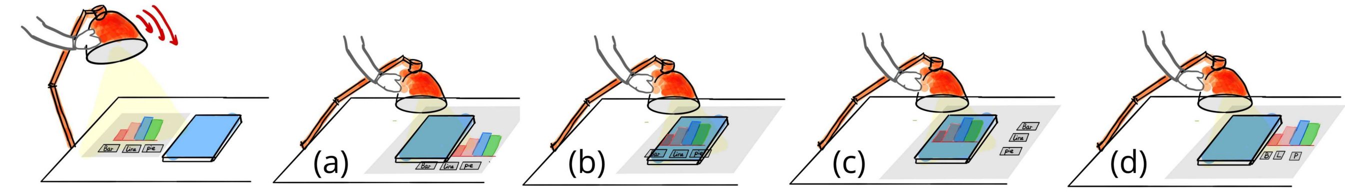


Figure 6.1: Adaptive content placement: (a) dodge (b) overlay (c) split & (d) shrink

This idea can be investigated through two projects. The first project would investigate factors that influence content placement and explore four interaction techniques to enable adaptive content placement (Figure 6.1): a) *dodge*: the virtual content avoids non-planar and obscured surfaces in order to self-accommodate in a clean area. b) *Overlay*: the virtual content is superimposed on non-planar surfaces. c) *split*: the virtual content will be divided into distinct components. d) *shrink*: the virtual information reduces the level of detail while retaining context. The second project is positioned over the previous project, investigating how the content should be further adapted to avoid occlusion induced by interactions from users, such as gestures, touch, and inputs from other devices, highlighting users' concerns, and understanding user preferences and behaviors.

Collaboration over Pervasive Desktop Computing

Another promising avenue for exploration is the concept of shared physical activities. This idea extends beyond simply sharing physical and digital objects on surfaces [117, 173], and leverages the surface adaptation capabilities of LuxAR and the ambient display concepts from Vogel and Balakrishnan's work [234], investigating how digital information and collaborative task states can be dynamically associated with physical locations, objects, and user activities. Consider a scenario in which multiple team members collaborate on a physical task using augmented lamps. It is interesting to explore how people use physical objects and surfaces as anchor points for virtual information. Team members could leave virtual notes, progress updates, or even partially completed tasks "attached" to relevant physical spaces or items. This would create a persistent, context-aware information layer that others can discover and interact with, fostering asynchronous collaboration and knowledge sharing. These shared information spaces could also improve the awareness of colleagues' activities. Unlike the "black window" commonly encountered in video conferencing, this approach could provide subtle, context-rich indicators of team members' focus areas, availability, or progress on shared tasks.

This idea can be investigated through the following steps. An initial study would investigate the potential of various physical objects and surfaces around the desktop as anchor points for conveying information in a collaborative environment. The research questions could include: 1) Which physical objects and surfaces are the most intuitive and effective for anchoring virtual information? 2) What types of information are best suited for different physical anchors? 3) How do users naturally interact with these augmented physical objects? We could conduct semi-structured interviews with think-aloud protocols. The results would provide a consensus on the usage of physical objects and surfaces as anchor points for virtual information. We can then develop a comprehensive framework to define the use of physical objects and surfaces as information anchors, including categories such as "information persistence", "interaction modality", "visibility range", and "update frequency". We hope this could guide future research and development in pervasive desktop computing for collaborative environments.

6.7 Final word

Desktop computing has long been, and continues to be, one of the dominant computing environment. Many efforts have been made to make desktop computing pervasive due to the potential benefits by merging the power and familiarity of traditional desktop systems with the flexibility and context-awareness inherent to pervasive computing. This thesis contributes to this endeavor through the use of LuxAR, an augmented lamp. This augmented lamp serves both as an input and output device, enabling the design of diverse interaction techniques through its direct manipulation. It not only creates a bidirectional connection across different computing environments, so that virtual information can be transferred between displays, devices, and surfaces, but also maintains an awareness of both the physical environment and users' activities, ensuring that the digital augmentations remain contextually relevant. Building on this system, the thesis investigates further, exploring how direct manipulation of the lamp opens up new possibilities for interaction. We investigate how coupled and decoupled control mechanisms impact target acquisition in a dynamic peephole display. Based on our observations, we examine a broader interaction context, focusing on the effects of intended use of a target on its initial acquisition. This exploration provides valuable insights into how users' intentions shape their behaviors with a virtual target. Collectively, these projects represent a step towards in bridging the gap between our digital and physical worlds, pointing towards a future where our digital and physical experiences are more entangled, and computing is not only conventional within a physical display, but spread out everywhere in the physical environment.

[1] Gregory D Abowd and Russell Beale. *"Users, systems and interfaces: A unifying framework for interaction."* In: *HCI*. Vol. 91. 1991, pp. 73–87.

[2] Gregory D. Abowd, Anind K. Dey, Peter J. Brown, Nigel Davies, Mark Smith, and Pete Steggles. *"Towards a Better Understanding of Context and Context-Awareness."* In: *Handheld and Ubiquitous Computing*. Ed. by Hans-W. Gellersen. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999, pp. 304–307. ISBN: 978-3-540-48157-7.

[3] Johnny Accot and Shumin Zhai. *"Beyond Fitts' Law: Models for Trajectory-Based HCI Tasks."* In: *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*. CHI '97. New York, NY, USA: Association for Computing Machinery, 1997, 295–302. ISBN: 0897918029. DOI: 10.1145/258549.258760. URL: https://doi.org/10.1145/258549.258760.

[4] Johnny Accot and Shumin Zhai. *"Refining Fitts' Law Models for Bivariate Pointing."* In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '03. New York, NY, USA: Association for Computing Machinery, 2003, 193–200. ISBN: 1581136307. DOI: 10.1145/642611.642646.

[5] Amani Alkayyali, Yasha Iravantchi, Jaylin Herskovitz, and Alanson P Sample. *"UbiChromics: Enabling Ubiquitously Deployable Interactive Displays with Photochromic Paint."* In: *Proceedings of the ACM on Human-Computer Interaction* 6.ISS (2022), pp. 118–142.

[6] Christopher Andrews, Alex Endert, and Chris North. *"Space to Think: Large High-Resolution Displays for Sensemaking."* In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '10. New York, NY, USA: Association for Computing Machinery, 2010, 55–64. ISBN: 9781605589299. DOI: 10.1145/1753326.1753336. URL: https://doi.org/10.1145/1753326.1753336.

[7] *Apple Vision Pro Full Specs, Features and Details*. 2024. URL: https://www.uploadvr.com/apple-vision-pro-specs/.

[8] *Apple Vision Pro Review: Tomorrow's Ideas... Today's Tech!* 2024. URL: https://www.youtube.com/watch?v=86Gy035z_KA/.

[9] Tomohiro Araki and Takashi Komuro. *"On-mouse projector: peephole interaction using a mouse with a projector."* In: *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia*. MUM '17. New York, NY, USA: Association for Computing Machinery, 2017, 231–239. ISBN: 9781450353786. DOI: 10.1145/3152832.3152849. URL: https://doi.org/10.1145/3152832.3152849.

[10] *Arduino MKR WiFi 1010*. 2024. URL: https://store-usa.arduino.cc/products/arduino-mkr-wifi-1010/.

[11] Rahul Arora, Rubaiat Habib Kazi, Danny M. Kaufman, Wilmot Li, and Karan Singh. *"MagicalHands: Mid-Air Hand Gestures for Animating in VR."* In: *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. UIST '19. New York, NY, USA: Association for Computing Machinery, 2019, 463–477. ISBN: 9781450368162. DOI: 10.1145/3332165.3347942. URL: https://doi.org/10.1145/3332165.3347942.

[12] Rahul Arora, Jiannan Li, Gongyi Shi, and Karan Singh. *Thinking Outside the Lab: VR Size & Depth Perception in the Wild*. 2021. arXiv: 2105.00584 [cs.HC].

[13] Mark Ashdown and Peter Robinson. *"A Personal Projected Display."* In: *Proceedings of the 12th Annual ACM International Conference on Multimedia*. MULTIMEDIA '04. New York, NY, USA: Association for Computing Machinery, 2004, 932–933. ISBN: 1581138938. DOI: 10.1145/1027527.1027739. URL: https://doi.org/10.1145/1027527.1027739.

[14] Gilles Bailly, Sidharth Sahdev, Sylvain Malacria, and Thomas Pietrzak. *"LivingDesktop: Augmenting Desktop Workstation with Actuated Devices."* In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. CHI '16. New York, NY, USA: Association for Computing Machinery, 2016, 5298–5310. ISBN: 9781450333627. DOI: 10.1145/2858036.2858208. URL: https://doi.org/10.1145/2858036.2858208.

[15] Ravin Balakrishnan. *""Beating" Fitts' law: virtual enhancements for pointing facilitation."* In: *International Journal of Human-Computer Studies* 61.6 (2004). Fitts' law 50 years later: applications and contributions from human-computer interaction, pp. 857–874. ISSN: 1071-5819. DOI: https://doi.org/10.1016/j.ijhcs.2004.09.002. URL: https://www.sciencedirect.com/science/article/pii/S107158190400103X.

[16] Mark S Baldwin, Gillian R Hayes, Oliver L Haimson, Jennifer Mankoff, and Scott E Hudson. *"The tangible desktop: a multimodal approach to nonvisual computing."* In: *ACM Transactions on Accessible Computing (TACCESS)* 10.3 (2017), pp. 1–28.

[17] 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*. New York, NY, USA: Association for Computing Machinery, 2019, 1–12. ISBN: 9781450359702. URL: https://doi.org/10.1145/3290605.3300331.

[18] 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. New York, NY, USA: Association for Computing Machinery, 2019, 1–12. ISBN: 9781450359702. DOI: 10.1145/3290605.3300331. URL: https://doi.org/10.1145/3290605.3300331.

[19] Guruduth Banavar, James Beck, Eugene Gluzberg, Jonathan Munson, Jeremy Sussman, and Deborra Zukowski. *"Challenges: an application model for pervasive computing."* In: *Proceedings of the 6th Annual International Conference on Mobile Computing and Networking*. MobiCom '00. New York, NY, USA: Association for Computing Machinery, 2000, 266–274. ISBN: 1581131976. DOI: 10.1145/345910.345957. URL: https://doi.org/10.1145/345910.345957.

[20] 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. New York, NY, USA: Association for Computing Machinery, 2019, 1–14. ISBN: 9781450359702. DOI: 10.1145/3290605.3300437. URL: https://doi.org/10.1145/3290605.3300437.

[21] Anil Ufuk Batmaz and Wolfgang Stuerzlinger. *"Effective Throughput Analysis of Different Task Execution Strategies for Mid-Air Fitts' Tasks in Virtual Reality."* In: *IEEE Transactions on Visualization and Computer Graphics* (2022), pp. 1–9. DOI: 10.1109/TVCG.2022.3203105.

[22] P. Beardsley, J. van Baar, R. Raskar, and C. Forlines. *"Interaction using a handheld projector."* In: *IEEE Computer Graphics and Applications* 25.1 (2005), pp. 39–43. DOI: 10.1109/MCG.2005.12.

[23] Hrvoje Benko. *"Beyond flat surface computing: challenges of depth-aware and curved interfaces."* In: *Proceedings of the 17th ACM international conference on Multimedia*. 2009, pp. 935–944.

[24] Hrvoje Benko, Andrew D. Wilson, and Federico Zannier. *"Dyadic Projected Spatial Augmented Reality."* In: *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*. UIST '14. New York, NY, USA: Association for Computing Machinery, 2014, 645–655. ISBN: 9781450330695. DOI: 10.1145/2642918.2647402. URL: https://doi.org/10.1145/2642918.2647402.

[25] Joanna Bergström, Tor-Salve Dalsgaard, Jason Alexander, and Kasper Hornbæk. *"How to Evaluate Object Selection and Manipulation in VR? Guidelines from 20 Years of Studies."* In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2021. ISBN: 9781450380966. URL: https://doi.org/10.1145/3411764.3445193.

[26] Xiaojun Bi and Ravin Balakrishnan. *"Comparing usage of a large high-resolution display to single or dual desktop displays for daily work."* In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2009, pp. 1005–1014.

[27] Xiaojun Bi, Tovi Grossman, Justin Matejka, and George Fitzmaurice. *"Magic Desk: Bringing Multi-Touch Surfaces into Desktop Work."* In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '11. New York, NY, USA: Association for Computing Machinery, 2011, 2511–2520. ISBN: 9781450302289. DOI: 10.1145/1978942.1979309. URL: https://doi.org/10.1145/1978942.1979309.

[28] Verena Biener et al. *"Quantifying the Effects of Working in VR for One Week."* In: *IEEE Transactions on Visualization and Computer Graphics* 28.11 (2022), pp. 3810–3820. DOI: 10.1109/TVCG.2022.3203103.

[29] Eric A. Bier, Maureen C. Stone, Ken Pier, William Buxton, and Tony D. DeRose. *"Toolglass and Magic Lenses: The See-through Interface."* In: *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques.* SIGGRAPH '93. New York, NY, USA: Association for Computing Machinery, 1993, 73–80. ISBN: 0897916018. DOI: 10.1145/166117.166126. URL: https://doi.org/10.1145/166117.166126.

[30] Eric A. Bier, Maureen C. Stone, Ken Pier, William Buxton, and Tony D. DeRose. *"Toolglass and Magic Lenses: The See-Through Interface."* In: *Seminal Graphics Papers: Pushing the Boundaries, Volume 2.* 1st ed. New York, NY, USA: Association for Computing Machinery, 2023. ISBN: 9798400708978. URL: https://doi.org/10.1145/3596711.3596719.

[31] Oliver Bimber and Ramesh Raskar. *Spatial augmented reality: merging real and virtual worlds.* CRC press, 2005.

[32] G. Blasko, F. Coriand, and S. Feiner. *"Exploring interaction with a simulated wrist-worn projection display."* In: *Ninth IEEE International Symposium on Wearable Computers (ISWC'05).* 2005, pp. 2–9. DOI: 10.1109/ISWC.2005.21.

[33] Florian Block, Hans Gellersen, and Nicolas Villar. *"Touch-Display Keyboards: Transforming Keyboards into Interactive Surfaces."* In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* CHI '10. New York, NY, USA: Association for Computing Machinery, 2010, 1145–1154. ISBN: 9781605589299. DOI: 10.1145/1753326.1753498. URL: https://doi.org/10.1145/1753326.1753498.

[34] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. *"Yolov4: Optimal speed and accuracy of object detection."* In: *arXiv preprint arXiv:2004.10934* (2020).

[35] 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. New York, NY, USA: Association for Computing Machinery, 2010, 2287–2296. ISBN: 9781605589299. DOI: 10.1145/1753326.1753671. URL: https://doi.org/10.1145/1753326.1753671.

[36] Riccardo Bovo, Eric J. Gonzalez, Li-Te Cheng, and Mar Gonzalez-Franco. *"WindowMirror: An Opensource Toolkit to Bring Interactive Multi-Window Views into XR."* In: *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW).* 2024, pp. 436–438. DOI: 10.1109/VRW62533.2024.00085.

[37] G. E. P. Box and D. R. Cox. *"An analysis of transformations."* In: *Journal of the Royal Statistical Society. Series B (Methodological* (1964), pp. 211–252.

[38] Anke M. Brock, Julia Chatain, Michelle Park, Tommy Fang, Martin Hachet, James A. Landay, and Jessica R. Cauchard. *"FlyMap: Interacting with Maps Projected from a Drone."* In: *Proceedings of the 7th ACM International Symposium on Pervasive Displays*. PerDis '18. New York, NY, USA: Association for Computing Machinery, 2018. ISBN: 9781450357654. DOI: 10.1145/3205873.3205877. URL: https://doi.org/10.1145/3205873.3205877.

[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. New York, NY, USA: Association for Computing Machinery, 2019, 1–28. ISBN: 9781450359702. DOI: 10.1145/3290605.3300792. URL: https://doi.org/10.1145/3290605.3300792.

[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. New York, NY, USA: Association for Computing Machinery, 2019. ISBN: 9781450368254. DOI: 10.1145/3338286.3340113. URL: https://doi.org/10.1145/3338286.3340113.

[41] 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 2017 ACM International Conference on Interactive Surfaces and Spaces*. 2017, pp. 62–71.

[42] Andreas Butz and Antonio Kruger. *"Applying the peephole metaphor in a mixed-reality room."* In: *IEEE Computer Graphics and applications* 26.1 (2006), pp. 56–63.

[43] François Bérard. *"Congruent Indirect Touch vs. mouse pointing performance."* In: *International Journal of Human-Computer Studies* 187 (2024), p. 103261. ISSN: 1071-5819. DOI: https://doi.org/10.1016/j.ijhcs.2024.103261. URL: https://www.sciencedirect.com/science/article/pii/S1071581924000454.

[44] Carlos Campos, Richard Elvira, Juan J. Gómez Rodríguez, José M. M. Montiel, and Juan D. Tardós. *"ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual–Inertial, and Multimap SLAM."* In: *IEEE Transactions on Robotics* 37.6 (2021), pp. 1874–1890. DOI: 10.1109/TRO.2021.3075644.

[45] Xiang Cao and Ravin Balakrishnan. *"Interacting with Dynamically Defined Information Spaces Using a Handheld Projector and a Pen."* In: *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology*. UIST '06. New York, NY, USA: Association for Computing

Machinery, 2006, 225–234. ISBN: 1595933131. DOI: 10.1145/1166253.1166289. URL: https://doi.org/10.1145/1166253.1166289.

[46] Xiang Cao, Clifton Forlines, and Ravin Balakrishnan. "*Multi-User Interaction Using Handheld Projectors*." In: *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology*. UIST '07. New York, NY, USA: Association for Computing Machinery, 2007, 43–52. ISBN: 9781595936790. DOI: 10.1145/1294211.1294220. URL: https://doi.org/10.1145/1294211.1294220.

[47] Xiang Cao, Jacky Jie Li, and Ravin Balakrishnan. "*Peephole Pointing: Modeling Acquisition of Dynamically Revealed Targets*." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '08. New York, NY, USA: Association for Computing Machinery, 2008, 1699–1708. ISBN: 9781605580111. DOI: 10.1145/1357054.1357320. URL: https://doi.org/10.1145/1357054.1357320.

[48] Xiang Cao, Jacky Jie Li, and Ravin Balakrishnan. "*Peephole Pointing: Modeling Acquisition of Dynamically Revealed Targets*." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '08. New York, NY, USA: Association for Computing Machinery, 2008, 1699–1708. ISBN: 9781605580111. DOI: 10.1145/1357054.1357320. URL: https://doi.org/10.1145/1357054.1357320.

[49] Stuart K. Card, Thomas P. Moran, and Allen Newell. "*The Keystroke-Level Model for User Performance Time with Interactive Systems*." In: *Commun. ACM* 23.7 (1980), 396–410. ISSN: 0001-0782. DOI: 10.1145/358886.358895. URL: https://doi.org/10.1145/358886.358895.

[50] 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. New York, NY, USA: Association for Computing Machinery, 2012, 2527–2530. ISBN: 9781450310154. DOI: 10.1145/2207676.2208639. URL: https://doi.org/10.1145/2207676.2208639.

[51] Jessica R. Cauchard, Mike Fraser, Teng Han, and Sriram Subramanian. "*Steerable Projection: Exploring Alignment in Interactive Mobile Displays*." In: *Personal Ubiquitous Comput.* 16.1 (2012), 27–37. ISSN: 1617-4909. DOI: 10.1007/s00779-011-0375-3. URL: https://doi.org/10.1007/s00779-011-0375-3.

[52] Jessica Cauchard, Markus Löchtefeld, Mike Fraser, Antonio Krüger, and Sriram Subramanian. "*m+ pSpaces: virtual workspaces in the spatially-aware mobile environment*." In: *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*. 2012, pp. 171–180.

[53] Paul E Ceruzzi. *A history of modern computing*. MIT press, 2003.

[54] Li-Wei Chan, Hsiang-Tao Wu, Hui-Shan Kao, Ju-Chun Ko, Home-Ru Lin, Mike Y. Chen, Jane Hsu, and Yi-Ping Hung. *"Enabling Beyond-Surface Interactions for Interactive Surface with an Invisible Projection."* In: *Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology*. UIST '10. New York, NY, USA: Association for Computing Machinery, 2010, 263–272. ISBN: 9781450302715. DOI: 10.1145/1866029.1866072. URL: https://doi.org/10.1145/1866029.1866072.

[55] Li-Wei Chan, Wei-Shian Ye, Shou-Chun Liao, Yu-Pao Tsai, Jane Hsu, and Yi-Ping Hung. *"A Flexible Display by Integrating a Wall-Size Display and Steerable Projectors."* In: *Ubiquitous Intelligence and Computing*. Ed. by Jianhua Ma, Hai Jin, Laurence T. Yang, and Jeffrey J.-P. Tsai. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 21–31. ISBN: 978-3-540-38092-4.

[56] Yuan Chen, Géry Casiez, Sylvain Malacria, and Daniel Vogel. *"LuxAR: A Direct Manipulation Projected Display to Extend and Augment Desktop Computing."* In: *In Graphics Interface (GI '24)*. GI '24. New York, NY, USA: Association for Computing Machinery, 2024, 1–12. ISBN: 9798400718281. DOI: 10.1145/3670947.3670981. URL: https://doi.org/10.1145/3670947.3670981.

[57] 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. New York, NY, USA: Association for Computing Machinery, 2020, 1–13. ISBN: 9781450367080. DOI: 10.1145/3313831.3376592. URL: https://doi.org/10.1145/3313831.3376592.

[58] Yuan Chen, Junwei Sun, Qiang Xu, Edward Lank, Pourang Irani, and Wei Li. *"Empirical Evaluation of Moving Target Selection in Virtual Reality Using Egocentric Metaphors."* In: *Human-Computer Interaction – INTERACT 2021*. Ed. by Carmelo Ardito, Rosa Lanzilotti, Alessio Malizia, Helen Petrie, Antonio Piccinno, Giuseppe Desolda, and Kori Inkpen. Cham: Springer International Publishing, 2021, pp. 29–50. ISBN: 978-3-030-85610-6.

[59] Yuan Chen, Junwei Sun, Qiang Xu, Edward Lank, Pourang Irani, and Wei Li. *"Global Scene Filtering, Exploration, and Pointing in Occluded Virtual Space."* In: *IFIP Conference on Human-Computer Interaction*. Springer. 2021, pp. 156–176.

[60] Kai-Yin Cheng, Rong-Hao Liang, Bing-Yu Chen, Rung-Huei Laing, and Sy-Yen Kuo. *"ICon: Utilizing Everyday Objects as Additional, Auxiliary and Instant Tabletop Controllers."* In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '10. New York, NY, USA: Association for Computing Machinery, 2010, 1155–1164. ISBN: 9781605589299. DOI: 10.1145/1753326.1753499. URL: https://doi.org/10.1145/1753326.1753499.

[61] Yifei Cheng, Yukang Yan, Xin Yi, Yuanchun Shi, and David Lindlbauer. "*Semanticadapt: Optimization-based adaptation of mixed reality layouts leveraging virtual-physical semantic connections*." In: *The 34th Annual ACM Symposium on User Interface Software and Technology*. 2021, pp. 282–297.

[62] Nalin Chhibber, Hemant Bhaskar Surale, Fabrice Matulic, and Daniel Vogel. "*Typealike: Near-keyboard hand postures for expanded laptop interaction*." In: *Proceedings of the ACM on Human-Computer Interaction* 5.ISS (2021), pp. 1–20.

[63] Andy Cockburn and Bruce McKenzie. "*Evaluating the Effectiveness of Spatial Memory in 2D and 3D Physical and Virtual Environments*." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '02. New York, NY, USA: Association for Computing Machinery, 2002, 203–210. ISBN: 1581134533. DOI: 10.1145/503376.503413. URL: https://doi.org/10.1145/503376.503413.

[64] Robert H Collins. "*Microcomputer Applications: Unleash the Power of Desktop Presentations*." In: *Journal of Personal Selling & Sales Management* 9.1 (1989), pp. 70–75.

[65] Christian Corsten, Ignacio Avellino, Max Möllers, and Jan Borchers. "*Instant User Interfaces: Repurposing Everyday Objects as Input Devices*." In: *Proceedings of the 2013 ACM International Conference on Interactive Tabletops and Surfaces*. ITS '13. New York, NY, USA: Association for Computing Machinery, 2013, 71–80. ISBN: 9781450322713. DOI: 10.1145/2512349.2512799. URL: https://doi.org/10.1145/2512349.2512799.

[66] Lisa G. Cowan and Kevin A. Li. "*ShadowPuppets: Supporting Collocated Interaction with Mobile Projector Phones Using Hand Shadows*." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '11. New York, NY, USA: Association for Computing Machinery, 2011, 2707–2716. ISBN: 9781450302289. DOI: 10.1145/1978942.1979340. URL: https://doi.org/10.1145/1978942.1979340.

[67] Raimund Dachselt and Sarmad Al-Saiegh. "*Interacting with printed books using digital pens and smart mobile projection*." In: *Proc. of the Workshop on Mobile and Personal Projection (MP2)@ ACM CHI*. Vol. 11. 2011.

[68] Kiran Dandekar, Balasundar I Raju, and Mandayam A Srinivasan. "*3-D finite-element models of human and monkey fingertips to investigate the mechanics of tactile sense*." In: *J. Biomech. Eng.* 125.5 (2003), pp. 682–691.

[69] Ciro Donalek, S George Djorgovski, Alex Cioc, Anwell Wang, Jerry Zhang, Elizabeth Lawler, Stacy Yeh, Ashish Mahabal, Matthew Graham, Andrew Drake, et al. "*Immersive and collaborative data visualization using virtual reality platforms*." In: *2014 IEEE International Conference on Big Data (Big Data)*. IEEE. 2014, pp. 609–614.

[70] Natalia Dużmańska, Paweł Strojny, and Agnieszka Strojny. "*Can simulator sickness be avoided? A review on temporal aspects of simulator sickness*." In: *Frontiers in psychology* 9 (2018), p. 410742.

[71] Karim El Batran and Mark D. Dunlop. *"Enhancing KLM (Keystroke-Level Model) to Fit Touch Screen Mobile Devices."* In: *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices & Services*. MobileHCI '14. New York, NY, USA: Association for Computing Machinery, 2014, 283–286. ISBN: 9781450330046. DOI: 10.1145/2628363.2628385. URL: https://doi.org/10.1145/2628363.2628385.

[72] Lisa A. Elkin, Matthew Kay, James J. Higgins, and Jacob O. Wobbrock. *"An Aligned Rank Transform Procedure for Multifactor Contrast Tests."* In: *The 34th Annual ACM Symposium on User Interface Software and Technology*. UIST '21. New York, NY, USA: Association for Computing Machinery, 2021, 754–768. ISBN: 9781450386357. DOI: 10.1145/3472749.3474784. URL: https://doi.org/10.1145/3472749.3474784.

[73] Satu Elo and Helvi Kyngäs. *"The qualitative content analysis process."* In: *Journal of advanced nursing* 62.1 (2008), pp. 107–115.

[74] Ahmed Elsharkawy, Khawar Naheem, Dongwoo Koo, and Mun Sang Kim. *"A UWB-driven self-actuated projector platform for interactive augmented reality applications."* In: *Applied Sciences* 11.6 (2021), p. 2871.

[75] Alex Endert, Lauren Bradel, Jessica Zeitz, Christopher Andrews, and Chris North. *"Designing Large High-Resolution Display Workspaces."* In: *Proceedings of the International Working Conference on Advanced Visual Interfaces*. AVI '12. New York, NY, USA: Association for Computing Machinery, 2012, 58–65. ISBN: 9781450312875. DOI: 10.1145/2254556.2254570. URL: https://doi.org/10.1145/2254556.2254570.

[76] Barrett Ens, David Ahlström, and Pourang Irani. *"Moving Ahead with Peephole Pointing: Modelling Object Selection with Head-Worn Display Field of View Limitations."* In: *Proceedings of the 2016 Symposium on Spatial User Interaction*. SUI '16. New York, NY, USA: Association for Computing Machinery, 2016, 107–110. ISBN: 9781450340687. DOI: 10.1145/2983310.2985756. URL: https://doi.org/10.1145/2983310.2985756.

[77] Caleb Escobedo, Matthew Strong, Mary West, Ander Aramburu, and Alessandro Roncone. *"Contact anticipation for physical human–robot interaction with robotic manipulators using onboard proximity sensors."* In: *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2021, pp. 7255–7262.

[78] *Exceptional PC VR Comfort With Significant Tradeoffs*. 2024. URL: https://www.uploadvr.com/bigscreen-beyond-hardware-review/.

[79] Jerry Alan Fails, Allison Druin, Mona Leigh Guha, Gene Chipman, Sante Simms, and Wayne Churaman. *"Child's play: a comparison of desktop and physical interactive environments."* In: *Proceedings of the 2005 conference on Interaction design and children*. 2005, pp. 48–55.

[80] Katherine Fennedy, Jeremy Hartmann, Quentin Roy, Simon T. Perrault, and Daniel Vogel. *"OctoPocus in VR: Using a Dynamic Guide for 3D Mid-Air Gestures in Virtual Reality."* In: *IEEE Transactions on Visualization and Computer Graphics* (2021), pp. 1–1. DOI: 10.1109/TVCG.2021.3101854.

[81] Cornelia Fermüller, Fang Wang, Yezhou Yang, Konstantinos Zampogiannis, Yi Zhang, Francisco Barranco, and Michael Pfeiffer. "*Prediction of manipulation actions*." In: *International Journal of Computer Vision* 126 (2018), pp. 358–374.

[82] P. M. Fitts. "*The information capacity of the human motor system in controlling the amplitude of movement*." In: *Journal of experimental psychology* 47 6 (1954), pp. 381–91.

[83] George W. Fitzmaurice. "*Situated Information Spaces and Spatially Aware Palmtop Computers*." In: *Commun. ACM* 36.7 (1993), 39–49. ISSN: 0001-0782. DOI: 10.1145/159544.159566. URL: https://doi.org/10.1145/159544.159566.

[84] George W. Fitzmaurice, Shumin Zhai, and Mark H. Chignell. "*Virtual Reality for Palmtop Computers*." In: *ACM Trans. Inf. Syst.* 11.3 (1993), 197–218. ISSN: 1046-8188. DOI: 10.1145/159161.159160. URL: https://doi.org/10.1145/159161.159160.

[85] Clifton Forlines, Ravin Balakrishnan, Paul Beardsley, Jeroen van Baar, and Ramesh Raskar. "*Zoom-and-Pick: Facilitating Visual Zooming and Precision Pointing with Interactive Handheld Projectors*." In: *Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology*. UIST '05. New York, NY, USA: Association for Computing Machinery, 2005, 73–82. ISBN: 1595932712. DOI: 10.1145/1095034.1095046. URL: https://doi.org/10.1145/1095034.1095046.

[86] Eisuke Fujinawa, Kenji Goto, Atsushi Irie, Songtao Wu, and Kuanhong Xu. "*Occlusion-aware hand posture based interaction on tabletop projector*." In: *Adjunct Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. 2019, pp. 113–115.

[87] Maurizio Gentilucci. "*Object motor representation and reaching–grasping control*." In: *Neuropsychologia* 40.8 (2002), pp. 1139–1153. ISSN: 0028-3932. DOI: https://doi.org/10.1016/S0028-3932(01)00233-0. URL: https://www.sciencedirect.com/science/article/pii/S0028393201002330.

[88] Renaud Gervais, Joan Sol Roo, and Martin Hachet. "*Tangible viewports: Getting out of flatland in desktop environments*." In: *Proceedings of the TEI'16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction*. 2016, pp. 176–184.

[89] Alix Goguey, Mathieu Nancel, Géry Casiez, and Daniel Vogel. "*The Performance and Preference of Different Fingers and Chords for Pointing, Dragging, and Object Transformation*." In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. CHI '16. New York, NY, USA: Association for Computing Machinery, 2016, 4250–4261. ISBN: 9781450333627. DOI: 10.1145/2858036.2858194. URL: https://doi.org/10.1145/2858036.2858194.

[90] Alix Goguey, Mathieu Nancel, Géry Casiez, and Daniel Vogel. *"The Performance and Preference of Different Fingers and Chords for Pointing, Dragging, and Object Transformation."* In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. CHI '16. New York, NY, USA: Association for Computing Machinery, 2016, 4250–4261. ISBN: 9781450333627. DOI: 10.1145/2858036.2858194. URL: https://doi.org/10.1145/2858036.2858194.

[91] *Google Earth VR*. 2022. URL: https://arvr.google.com/earth/.

[92] Evan D. Graham and Christine L. MacKenzie. *"Physical versus Virtual Pointing."* In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '96. New York, NY, USA: Association for Computing Machinery, 1996, 292–299. ISBN: 0897917774. DOI: 10.1145/238386.238532. URL: https://doi.org/10.1145/238386.238532.

[93] Andrew Greaves and Enrico Rukzio. *"Evaluation of Picture Browsing Using a Projector Phone."* In: *Proceedings of the 10th International Conference on Human Computer Interaction with Mobile Devices and Services*. MobileHCI '08. New York, NY, USA: Association for Computing Machinery, 2008, 351–354. ISBN: 9781595939524. DOI: 10.1145/1409240.1409286. URL: https://doi.org/10.1145/1409240.1409286.

[94] Jonathan Grudin. *"Partitioning digital worlds: focal and peripheral awareness in multiple monitor use."* In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '01. New York, NY, USA: Association for Computing Machinery, 2001, 458–465. ISBN: 1581133278. DOI: 10.1145/365024.365312. URL: https://doi.org/10.1145/365024.365312.

[95] Yves Guiard. *"Asymmetric Division of Labor in Human Skilled Bimanual Action."* In: *Journal of Motor Behavior* 19.4 (1987). PMID: 15136274, pp. 486–517. DOI: 10.1080/00222895.1987.10735426. eprint: https://doi.org/10.1080/00222895.1987.10735426. URL: https://doi.org/10.1080/00222895.1987.10735426.

[96] ERROL R. HOFFMANN. *"A comparison of hand and foot movement times."* In: *Ergonomics* 34.4 (1991). PMID: 1860460, pp. 397–406. DOI: 10.1080/00140139108967324. eprint: https://doi.org/10.1080/00140139108967324. URL: https://doi.org/10.1080/00140139108967324.

[97] Thomas Haigh and Paul E Ceruzzi. *A new history of modern computing*. MIT Press, 2021.

[98] *Half-life: Alyx*. 2022. URL: https://www.half-life.com/en/alyx/.

[99] Violet Yinuo Han, Hyunsung Cho, Kiyosu Maeda, Alexandra Ion, and David Lindlbauer. *"BlendMR: A Computational Method to Create Ambient Mixed Reality Interfaces."* In: *Proceedings of the ACM on Human-Computer Interaction* 7.ISS (2023), pp. 217–241.

[100] Alina Hang, Enrico Rukzio, and Andrew Greaves. *"Projector Phone: A Study of Using Mobile Phones with Integrated Projector for Interaction with Maps."* In: *Proceedings of the 10th International Conference on Human Computer Interaction with Mobile Devices and Services*. MobileHCI '08. New York, NY, USA: Association for Computing Machinery, 2008, 207–216. ISBN: 9781595939524. DOI: 10.1145/1409240.1409263. URL: https://doi.org/10.1145/1409240.1409263.

[101] John Hardy. *"Experiences: A Year in the Life of an Interactive Desk."* In: *Proceedings of the Designing Interactive Systems Conference*. DIS '12. New York, NY, USA: Association for Computing Machinery, 2012, 679–688. ISBN: 9781450312103. DOI: 10.1145/2317956.2318058. URL: https://doi.org/10.1145/2317956.2318058.

[102] Sandra G Hart. *"NASA-task load index (NASA-TLX); 20 years later."* In: *Proceedings of the human factors and ergonomics society annual meeting*. Vol. 50. 9. Sage publications Sage CA: Los Angeles, CA. 2006, pp. 904–908.

[103] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. *"Mask r-cnn."* In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2961–2969.

[104] Paul Hine, Loza Tadesse Mamo, and Narcis Pares. *"AR Magic Lantern: Group-Based Co-Located Augmentation Based on the World-as-Support AR Paradigm."* In: *Extended Abstracts of the 2022 CHI Conf. on Human Factors in Computing Systems*. CHI EA '22. New York, NY, USA: ACM, 2022. ISBN: 9781450391566. DOI: 10.1145/3491101.3519918. URL: https://doi.org/10.1145/3491101.3519918.

[105] Steven Houben, Paolo Tell, and Jakob E. Bardram. *"ActivitySpace: Managing Device Ecologies in an Activity-Centric Configuration Space."* In: *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces*. ITS '14. New York, NY, USA: Association for Computing Machinery, 2014, 119–128. ISBN: 9781450325875. DOI: 10.1145/2669485.2669493. URL: https://doi.org/10.1145/2669485.2669493.

[106] Sebastian Hubenschmid, Johannes Zagermann, Daniel Leicht, Harald Reiterer, and Tiare Feuchtner. *"ARound the Smartphone: Investigating the Effects of Virtually-Extended Display Size on Spatial Memory."* In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. CHI '23. New York, NY, USA: Association for Computing Machinery, 2023. ISBN: 9781450394215. DOI: 10.1145/3544548.3581438. URL: https://doi.org/10.1145/3544548.3581438.

[107] Jochen Huber, Jürgen Steimle, and Max Mühlhäuser. *"A model of embodied dynamic peephole pointing for hidden targets."* In: *Proceedings of HCI 2011 The 25th BCS Conference on Human Computer Interaction*. BCS Learning & Development. 2011.

[108] Wolfgang Hürst and Tair Bilyalov. *"Dynamic versus Static Peephole Navigation of VR Panoramas on Handheld Devices."* In: *Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia*. MUM '10. New York, NY, USA: Association for Computing Machinery, 2010. ISBN: 9781450304245. DOI: 10.1145/1899475.1899500. URL: https://doi.org/10.1145/1899475.1899500.

[109] Dugald Ralph Hutchings and John T Stasko. *"Revisiting Display Space Management: Understanding Current Practice to Inform Next-generation Design."* In: *Graphics Interface*. 2004, pp. 127–134.

[110] Hiroshi Ishii. *"The tangible user interface and its evolution."* In: *Communications of the ACM* 51.6 (2008), pp. 32–36.

[111] Hiroshi Ishii and Brygg Ullmer. *"Tangible bits: towards seamless interfaces between people, bits and atoms."* In: *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*. 1997, pp. 234–241.

[112] Thomas E. Jerde, John F. Soechting, and Martha Flanders. *"Coarticulation in Fluent Fingerspelling."* In: *Journal of Neuroscience* 23.6 (2003), pp. 2383–2393. ISSN: 0270-6474. DOI: 10.1523/JNEUROSCI.23-06-02383.2003. eprint: https://www.jneurosci.org/content/23/6/2383.full.pdf. URL: https://www.jneurosci.org/content/23/6/2383.

[113] Roland S Johansson and J Randall Flanagan. *"Coding and use of tactile signals from the fingertips in object manipulation tasks."* In: *Nature Reviews Neuroscience* 10.5 (2009), pp. 345–359.

[114] Brett R. Jones, Hrvoje Benko, Eyal Ofek, and Andrew D. Wilson. *"IllumiRoom: Peripheral Projected Illusions for Interactive Experiences."* In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '13. New York, NY, USA: Association for Computing Machinery, 2013, 869–878. ISBN: 9781450318990. DOI: 10.1145/2470654.2466112. URL: https://doi.org/10.1145/2470654.2466112.

[115] Brett Jones, Rajinder Sodhi, Michael Murdock, Ravish Mehra, Hrvoje Benko, Andrew Wilson, Eyal Ofek, Blair MacIntyre, Nikunj Raghuvanshi, and Lior Shapira. *"RoomAlive: Magical Experiences Enabled by Scalable, Adaptive Projector-camera Units."* In: *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*. UIST '14. New York, NY, USA: ACM, 2014, pp. 637–644. ISBN: 978-1-4503-3069-5. DOI: 10.1145/2642918.2647383. URL: http://doi.acm.org/10.1145/2642918.2647383.

[116] Nikhita Joshi, Antony Albert Raj Irudayaraj, Jeremy Hartmann, and Daniel Vogel. *"An Instrumented Office Chair with a Steerable Projector for Personal Spatial Augmented Reality."* In: *Symposium on Spatial User Interaction*. SUI '22. New York, NY, USA: Association for Computing Machinery, 2022. ISBN: 9781450399487. DOI: 10.1145/3565970.3567705. URL: https://doi.org/10.1145/3565970.3567705.

[117] Sasa Junuzovic, Kori Inkpen, Tom Blank, and Anoop Gupta. *"IllumiShare: Sharing Any Surface."* In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '12. New York, NY, USA: Association for Computing Machinery, 2012, 1919–1928. ISBN: 9781450310154. DOI: 10.1145/2207676.2208333. URL: https://doi.org/10.1145/2207676.2208333.

[118] Yasuhiro Kajiwara, Keisuke Tajimi, Keiji Uemura, Nobuchika Sakata, and Shogo Nishida. *"ClippingLight: A Method for Easy Snapshots with Projection Viewfinder and Tilt-Based Zoom Control."* In: *Proceedings of the 2nd Augmented Human International Conference*. AH '11. New York, NY, USA: Association for Computing Machinery, 2011. ISBN: 9781450304269. DOI: 10.1145/1959826.1959840. URL: https://doi.org/10.1145/1959826.1959840.

[119] Shaun K. Kane, Daniel Avrahami, Jacob O. Wobbrock, Beverly Harrison, Adam D. Rea, Matthai Philipose, and Anthony LaMarca. *"Bonfire: A Nomadic System for Hybrid Laptop-Tabletop Interaction."* In: *Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology*. UIST '09. New York, NY, USA: Association for Computing Machinery, 2009, 129–138. ISBN: 9781605587455. DOI: 10.1145/1622176.1622202. URL: https://doi.org/10.1145/1622176.1622202.

[120] Bonifaz Kaufmann and David Ahlström. *"Revisiting Peephole Pointing: A Study of Target Acquisition with a Handheld Projector."* In: *Proceedings of the 14th International Conference on Human-Computer Interaction with Mobile Devices and Services*. MobileHCI '12. New York, NY, USA: Association for Computing Machinery, 2012, 211–220. ISBN: 9781450311052. DOI: 10.1145/2371574.2371607. URL: https://doi.org/10.1145/2371574.2371607.

[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. New York, NY, USA: Association for Computing Machinery, 2013, 3173–3176. ISBN: 9781450318990. DOI: 10.1145/2470654.2466434. URL: https://doi.org/10.1145/2470654.2466434.

[122] Frederic Kerber, Antonio Krüger, and Markus Löchtefeld. *"Investigating the Effectiveness of Peephole Interaction for Smartwatches in a Map Navigation Task."* In: *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices & Services*. MobileHCI '14. New York, NY, USA: Association for Computing Machinery, 2014, 291–294. ISBN: 9781450330046. DOI: 10.1145/2628363.2628393. URL: https://doi.org/10.1145/2628363.2628393.

[123] Daekun Kim, Nikhita Joshi, and Daniel Vogel. *"Perspective and Geometry Approaches to Mouse Cursor Control in Spatial Augmented Reality."* In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. CHI '23. New York, NY, USA: Association for Computing

Machinery, 2023. ISBN: 9781450394215. DOI: 10.1145/3544548.3580849. URL: https://doi.org/10.1145/3544548.3580849.

[124] Jeongyun Kim, Jonghoon Seo, and Tack-Don Han. *"AR Lamp: Interactions on Projection-Based Augmented Reality for Interactive Learning."* In: *Proceedings of the 19th International Conference on Intelligent User Interfaces*. IUI '14. New York, NY, USA: Association for Computing Machinery, 2014, 353–358. ISBN: 9781450321846. DOI: 10.1145/2557500.2557505. URL: https://doi.org/10.1145/2557500.2557505.

[125] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. *"Segment anything."* In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023, pp. 4015–4026.

[126] Konstantin Klamka and Raimund Dachselt. *"IllumiPaper: Illuminated interactive paper."* In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 2017, pp. 5605–5618.

[127] N. Klopcar and J. Lenarcic. *"Kinematic Model for Determination of Human Arm Reachable Workspace."* In: *Meccanica* 40 (2005), pp. 203–219.

[128] Pascal Knierim, Thomas Kosch, and Albrecht Schmidt. *"The nomadic office: A location independent workspace through mixed reality."* In: *IEEE Pervasive Computing* 20.4 (2021), pp. 71–78.

[129] Pascal Knierim, Valentin Schwind, Anna Maria Feit, Florian Nieuwenhuizen, and Niels Henze. *"Physical Keyboards in Virtual Reality: Analysis of Typing Performance and Effects of Avatar Hands."* In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18. New York, NY, USA: Association for Computing Machinery, 2018, 1–9. ISBN: 9781450356206. DOI: 10.1145/3173574.3173919. URL: https://doi.org/10.1145/3173574.3173919.

[130] Hideki Koike, Yoichi Sato, and Yoshinori Kobayashi. *"Integrating paper and digital information on EnhancedDesk: a method for realtime finger tracking on an augmented desk system."* In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 8.4 (2001), pp. 307–322.

[131] Paul G Kry, Adeline Pihuit, Adrien Bernhardt, and Marie-Paule Cani. *"Handnavigator: Hands-on interaction for desktop virtual reality."* In: *Proceedings of the 2008 ACM symposium on Virtual reality software and technology*. 2008, pp. 53–60.

[132] Gregorij Kurillo, Alic Chen, Ruzena Bajcsy, and Jay Han. *"Evaluation of upper extremity reachable workspace using Kinect camera."* In: *Technology and health care : official journal of the European Society for Engineering and Medicine* 21 (2013). DOI: 10.3233/THC-130764.

[133] 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*. 2018, pp. 1–14.

[134] Edward Lank, Yi-Chun Nikko Cheng, and Jaime Ruiz. "*Endpoint Prediction Using Motion Kinematics*." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '07. New York, NY, USA: Association for Computing Machinery, 2007, 637–646. ISBN: 9781595935939. DOI: 10.1145/1240624.1240724. URL: https://doi.org/10.1145/1240624.1240724.

[135] *Lantern*. 2018. URL: https://nordprojects.co/lantern/.

[136] Lan Li, Fei Yu, Dongquan Shi, Jianping Shi, Zongjun Tian, Jiquan Yang, Xingsong Wang, and Qing Jiang. "*Application of virtual reality technology in clinical medicine*." In: *American journal of translational research* 9.9 (2017), p. 3867.

[137] Zhen Li, Michelle Annett, Ken Hinckley, Karan Singh, and Daniel Wigdor. "*HoloDoc: Enabling Mixed Reality Workspaces That Harness Physical and Digital Content*." In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. New York, NY, USA: Association for Computing Machinery, 2019, 1–14. ISBN: 9781450359702. DOI: 10.1145/3290605.3300917. URL: https://doi.org/10.1145/3290605.3300917.

[138] Chunyuan Liao, Hao Tang, Qiong Liu, Patrick Chiu, and Francine Chen. "*FACT: Fine-Grained Cross-Media Interaction with Documents via a Portable Hybrid Paper-Laptop Interface*." In: *Proceedings of the 18th ACM International Conference on Multimedia*. MM '10. New York, NY, USA: Association for Computing Machinery, 2010, 361–370. ISBN: 9781605589336. DOI: 10.1145/1873951.1874001. URL: https://doi.org/10.1145/1873951.1874001.

[139] Natan Linder and Pattie Maes. "*LuminAR: Portable Robotic Augmented Reality Interface Design and Prototype*." In: *Adjunct Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology*. UIST '10. New York, NY, USA: Association for Computing Machinery, 2010, 395–396. ISBN: 9781450304627. DOI: 10.1145/1866218.1866237. URL: https://doi.org/10.1145/1866218.1866237.

[140] David Lindlbauer, Anna Maria Feit, and Otmar Hilliges. "*Context-aware online adaptation of mixed reality interfaces*." In: *Proceedings of the 32nd annual ACM symposium on user interface software and technology*. 2019, pp. 147–160.

[141] Lee Lisle, Kylie Davidson, Leonardo Pavanatto, Ibrahim A Tahmid, Chris North, and Doug A Bowman. "*Spaces to Think: A Comparison of Small, Large, and Immersive Displays for the Sensemaking Process*." In: *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE. 2023, pp. 1084–1093.

[142] Chuanyi Liu and Rui Zhao. "*Find the 'Lost' Cursor: A Comparative Experiment of Visually Enhanced Cursor Techniques*." In: *Intelligent Computing Theories and Application*. Ed. by De-Shuang Huang, Kang-Hyun Jo, and Xiao-Long Zhang. Cham: Springer International Publishing, 2018, pp. 85–92. ISBN: 978-3-319-95933-7.

[143] Paul Lubos, Gerd Bruder, and Frank Steinicke. *"Analysis of direct selection in head-mounted display environments."* In: *2014 IEEE Symposium on 3D User Interfaces (3DUI)*. 2014, pp. 11–18. DOI: 10.1109/3DUI.2014.6798834.

[144] Kris Luyten, Kristof Verpoorten, and Karin Coninx. *"Ad-hoc co-located collaborative work with mobile devices."* In: *Proceedings of the 9th international conference on Human computer interaction with mobile devices and services*. 2007, pp. 507–514.

[145] I. Scott MacKenzie. *"A Note on the Information-Theoretic Basis for Fitts' Law."* In: *Journal of Motor Behavior* 21.3 (1989). PMID: 15136269, pp. 323–330. DOI: 10.1080/00222895.1989.10735486. eprint: https://doi.org/10.1080/00222895.1989.10735486. URL: https://doi.org/10.1080/00222895.1989.10735486.

[146] I. Scott MacKenzie. *"Fitts' Law as a Research and Design Tool in Human-Computer Interaction."* In: *Human-Computer Interaction* 7.1 (1992), pp. 91–139. DOI: 10.1207/s15327051hci0701\$_3.

[147] Wendy E Mackay. *"Augmented reality: linking real and virtual worlds: a new paradigm for interacting with computers."* In: *Proceedings of the working conference on Advanced visual interfaces*. 1998, pp. 13–21.

[148] Kosuke Maeda, Mitski Piekenbrock, Toshiki Sato, and Hideki Koike. *"A Tabletop System Using an OmniDirectional Projector-Camera."* In: *Proc. of the 2018 ACM Int. Conf. on Interactive Surfaces and Spaces*. ISS '18. New York, NY, USA: ACM, 2018, 311–314. ISBN: 9781450356947. DOI: 10.1145/3279778.3279911. URL: https://doi.org/10.1145/3279778.3279911.

[149] Regan L. Mandryk and Calvin Lough. *"The Effects of Intended Use on Target Acquisition."* In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2011, 1649–1652. ISBN: 9781450302289. URL: https://doi.org/10.1145/1978942.1979182.

[150] Damien Masson, Alix Goguey, Sylvain Malacria, and Géry Casiez. *"WhichFingers: Identifying Fingers on Touch Surfaces and Keyboards using Vibration Sensors."* In: *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*. UIST '17. New York, NY, USA: Association for Computing Machinery, 2017, 41–48. ISBN: 9781450349819. DOI: 10.1145/3126594.3126619. URL: https://doi.org/10.1145/3126594.3126619.

[151] Fabrice Matulic, Wolfgang Büschel, Michael Ying Yang, Stephan Ihrke, Anmol Ramraika, Carsten Rother, and Raimund Dachselt. *"Smart Ubiquitous Projection: Discovering Surfaces for the Projection of Adaptive Content."* In: *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. CHI EA '16. New York, NY, USA: Association for Computing Machinery, 2016, 2592–2600. ISBN: 9781450340823. DOI: 10.1145/2851581.2892545. URL: https://doi.org/10.1145/2851581.2892545.

[152] Jess McIntosh, Mike Fraser, Paul Worgan, and Asier Marzo. *"DeskWave: Desktop interactions using low-cost microwave doppler arrays."* In: *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 2017, pp. 1885–1892.

[153] Mark Mcgill, Aidan Kehoe, Euan Freeman, and Stephen Brewster. *"Expanding the Bounds of Seated Virtual Workspaces."* In: *ACM Trans. Comput.-Hum. Interact.* 27.3 (2020). ISSN: 1073-0516. DOI: 10.1145/3380959. URL: https://doi.org/10.1145/3380959.

[154] *MeetinVR - Business Meetings & Collaboration in VR*. 2022. URL: https://www.meetinvr.com.

[155] Sumit Mehra, Peter Werkhoven, and Marcel Worring. *"Navigating on Handheld Displays: Dynamic versus Static Peephole Navigation."* In: *ACM Trans. Comput.-Hum. Interact.* 13.4 (2006), 448–457. ISSN: 1073-0516. DOI: 10.1145/1188816.1188818. URL: https://doi.org/10.1145/1188816.1188818.

[156] David E Meyer, Richard A Abrams, Sylvan Kornblum, Charles E Wright, and JE Keith Smith. *"Optimality in human motor performance: ideal control of rapid aimed movements."* In: *Psychological review* 95.3 (1988), p. 340.

[157] *Minisense 100 Vibration sensors*. 2024. URL: https://www.te.com/en/product-CAT-PFS0011.html.

[158] Pranav Mistry, Pattie Maes, and Liyan Chang. *"WUW - Wear Ur World: A Wearable Gestural Interface."* In: *CHI '09 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '09. New York, NY, USA: Association for Computing Machinery, 2009, 4111–4116. ISBN: 9781605582474. DOI: 10.1145/1520340.1520626. URL: https://doi.org/10.1145/1520340.1520626.

[159] David Molyneaux, Shahram Izadi, David Kim, Otmar Hilliges, Steve Hodges, Xiang Cao, Alex Butler, and Hans Gellersen. *"Interactive environment-aware handheld projectors for pervasive computing spaces."* In: *Pervasive Computing: 10th International Conference, Pervasive 2012, Newcastle, UK, June 18-22, 2012. Proceedings 10*. Springer. 2012, pp. 197–215.

[160] Rebecca Moody. *Screen Time Statistics: Average Screen Time by Country*. 2024. URL: https://www.comparitech.com/tv-streaming/screen-time-statistics/.

[161] *Movable Panels for Multitasking and Creators Come to Horizon Feed*. 2024. URL: https://www.meta.com/blog/quest/meta-quest-v67-update-new-window-layout-creator-content-horizon-feed/.

[162] Jens Müller, Roman Rädle, Hans-Christian Jetter, and Harald Reiterer. *"An Experimental Comparison of Vertical and Horizontal Dynamic Peephole Navigation."* In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI '15. New York, NY, USA: Association for Computing Machinery, 2015, 1523–1526. ISBN: 9781450331456. DOI:

10.1145/2702123.2702227. URL: https://doi.org/10.1145/2702123.2702227.

[163] Jens Müller, Tobias Schwarz, Simon Butscher, and Harald Reiterer. "*Back to tangibility: a post-WIMP perspective on control room design*." In: *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces*. 2014, pp. 57–64.

[164] Martin Mundt and Tintu Mathew. "*An Evaluation of Pie Menus for System Control in Virtual Reality*." In: *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*. New York, NY, USA: Association for Computing Machinery, 2020. ISBN: 9781450375795. URL: https://doi.org/10.1145/3419249.3420146.

[165] Atsuo Murata. "*Extending effective target width in Fitts' law to a two-dimensional pointing task*." In: *International journal of human-computer interaction* 11.2 (1999), pp. 137–152.

[166] Mathieu Nancel and Edward Lank. "*Modeling User Performance on Curved Constrained Paths*." In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2017, 244–254. ISBN: 9781450346559. URL: https://doi.org/10.1145/3025453.3025951.

[167] William Newman and Pierre Wellner. "*A Desk Supporting Computer-Based Interaction with Paper Documents*." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '92. New York, NY, USA: Association for Computing Machinery, 1992, 587–592. ISBN: 0897915135. DOI: 10.1145/142750.143007. URL: https://doi.org/10.1145/142750.143007.

[168] Aziz Niyazov, Barrett Ens, Kadek Ananta Satriadi, Nicolas Mellado, Loïc Barthe, Tim Dwyer, and Marcos Serrano. "*User-driven constraints for layout optimisation in augmented reality*." In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 2023, pp. 1–16.

[169] Aziz Niyazov, Nicolas Mellado, Loic Barthe, and Marcos Serrano. "*Dynamic Decals: Pervasive Freeform Interfaces Using Constrained Deformable Graphical Elements*." In: *Proc. ACM Hum.-Comput. Interact.* 5.ISS (2021). DOI: 10.1145/3488538. URL: https://doi.org/10.1145/3488538.

[170] Donald Norman. "*Affordance, conventions, and design*." In: *Interactions* 6 (1999), pp. 38–42. DOI: 10.1145/301153.301168.

[171] Biswaksen Patnaik, Huaishu Peng, and Niklas Elmqvist. "*VisTorch: Interacting with Situated Visualizations using Handheld Projectors*." In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 2024, pp. 1–13.

[172] Leonardo Pavanatto, Chris North, Doug A. Bowman, Carmen Badea, and Richard Stoakley. *"Do we still need physical monitors? An evaluation of the usability of AR virtual monitors for productivity work."* In: *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. 2021, pp. 759–767. DOI: 10.1109/VR50410.2021.00103.

[173] Tomislav Pejsa, Julian Kantor, Hrvoje Benko, Eyal Ofek, and Andrew Wilson. *"Room2Room: Enabling Life-Size Telepresence in a Projected Augmented Reality Environment."* In: *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. CSCW '16. New York, NY, USA: Association for Computing Machinery, 2016, 1716–1725. ISBN: 9781450335928. DOI: 10.1145/2818048.2819965. URL: https://doi.org/10.1145/2818048.2819965.

[174] Gary Perelman, Marcos Serrano, Mathieu Raynal, Celia Picard, Mustapha Derras, and Emmanuel Dubois. *"The roly-poly mouse: Designing a rolling input device unifying 2d and 3d interaction."* In: *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 2015, pp. 327–336.

[175] Tu-Hoa Pham, Nikolaos Kyriazis, Antonis A Argyros, and Abderrahmane Kheddar. *"Hand-object contact force estimation from markerless visual tracking."* In: *IEEE transactions on pattern analysis and machine intelligence* 40.12 (2017), pp. 2883–2896.

[176] Stephen Pheasant and Christine Haslegrave. *Bodyspace: Anthropometry, Ergonomics and the Design of Work*. 2018. ISBN: 9781315375212. DOI: 10.1201/9781315375212.

[177] Claudio Pinhanez. *"The Everywhere Displays Projector: A Device to Create Ubiquitous Graphical Interfaces."* In: *Ubicomp 2001: Ubiquitous Computing*. Ed. by Gregory D. Abowd, Barry Brumitt, and Steven Shafer. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001, pp. 315–331. ISBN: 978-3-540-45427-4.

[178] I. Poupyrev, T. Ichikawa, S. Weghorst, and M. Billinghurst. *"Egocentric Object Manipulation in Virtual Environments: Empirical Evaluation of Interaction Techniques."* In: *Computer Graphics Forum* 17.3 (1998), pp. 41–52. DOI: 10.1111/1467-8659.00252.

[179] Ivan Poupyrev, Mark Billinghurst, Suzanne Weghorst, and Tadao Ichikawa. *"The Go-Go Interaction Technique: Non-Linear Mapping for Direct Manipulation in VR."* In: *Proceedings of the 9th Annual ACM Symposium on User Interface Software and Technology*. UIST '96. New York, NY, USA: Association for Computing Machinery, 1996, 79–80. ISBN: 0897917987. DOI: 10.1145/237091.237102.

[180] Kathrin Probst, David Lindlbauer, Patrick Greindl, Markus Trapp, Michael Haller, Bernhard Schwartz, and Andreas Schrempf. *"Rotating, tilting, bouncing: using an interactive chair to promote activity in office environments."* In: *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. 2013, pp. 79–84.

[181] Kathrin Probst, David Lindlbauer, Michael Haller, Bernhard Schwartz, and Andreas Schrempf. *"A chair as ubiquitous input device: exploring semaphoric chair gestures for focused and peripheral interaction."* In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2014, pp. 4097–4106.

[182] *ProjectFLX*. 2021. URL: https://lightform.com/lfx.

[183] 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. New York, NY, USA: Association for Computing Machinery, 2013, 23–32. ISBN: 9781450322713. DOI: 10.1145/2512349.2512811. URL: https://doi.org/10.1145/2512349.2512811.

[184] 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. New York, NY, USA: Association for Computing Machinery, 2014, 45–54. ISBN: 9781450325875. DOI: 10.1145/2669485.2669500. URL: https://doi.org/10.1145/2669485.2669500.

[185] Roman Rädle, Hans-Christian Jetter, Jens Müller, and Harald Reiterer. *"Bigger is Not Always Better: Display Size, Performance, and Task Load during Peephole Map Navigation."* In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '14. New York, NY, USA: Association for Computing Machinery, 2014, 4127–4136. ISBN: 9781450324731. DOI: 10.1145/2556288.2557071. URL: https://doi.org/10.1145/2556288.2557071.

[186] Stefan Rapp, G Michelitsch, M Osen, J Williams, M Barbisch, R Bohan, Z Valsan, and M Emele. *"SPOTLIGHT NAVIGATION: INTERACTION WITH A HANDHELD PROJECTION DEVICE."* In: 2004.

[187] Umar Rashid, Miguel A Nacenta, and Aaron Quigley. *"The cost of display switching: a comparison of mobile, large display and hybrid UI configurations."* In: *Proceedings of the International Working Conference on Advanced Visual Interfaces*. 2012, pp. 99–106.

[188] Ramesh Raskar, Jeroen van Baar, Paul Beardsley, Thomas Willwacher, Srinivas Rao, and Clifton Forlines. *"ILamps: Geometrically Aware and Self-Configuring Projectors."* In: *ACM Trans. Graph.* 22.3 (2003), 809–818. ISSN: 0730-0301. DOI: 10.1145/882262.882349. URL: https://doi.org/10.1145/882262.882349.

[189] Ramesh Raskar, Paul Beardsley, Jeroen van Baar, Yao Wang, Paul Dietz, Johnny Lee, Darren Leigh, and Thomas Willwacher. *"RFIG Lamps: Interacting with a Self-Describing World via Photosensing Wireless Tags and Projectors."* In: *ACM Trans. Graph.* 23.3 (2004), 406–415. ISSN: 0730-0301. DOI: 10.1145/1015706.1015738. URL: https://doi.org/10.1145/1015706.1015738.

[190] Ramesh Raskar, Greg Welch, Matt Cutts, Adam Lake, Lev Stesin, and Henry Fuchs. "*The Office of the Future: A Unified Approach to Image-Based Modeling and Spatially Immersive Displays*." In: *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '98. New York, NY, USA: Association for Computing Machinery, 1998, 179–188. ISBN: 0897919998. DOI: 10.1145/280814.280861. URL: https://doi.org/10.1145/280814.280861.

[191] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. "*You only look once: Unified, real-time object detection*." In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 779–788.

[192] Jun Rekimoto and Masanori Saitoh. "*Augmented Surfaces: A Spatially Continuous Work Space for Hybrid Computing Environments*." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '99. New York, NY, USA: Association for Computing Machinery, 1999, 378–385. ISBN: 0201485591. DOI: 10.1145/302979.303113. URL: https://doi.org/10.1145/302979.303113.

[193] Brett Ridel, Patrick Reuter, Jeremy Laviole, Nicolas Mellado, Nadine Couture, and Xavier Granier. "*The revealing flashlight: Interactive spatial augmented reality for detail exploration of cultural heritage artifacts*." In: *Journal on Computing and Cultural Heritage (JOCCH)* 7.2 (2014), pp. 1–18.

[194] Jan Riemann, Martin Schmitz, Alexander Hendrich, and Max Mühlhäuser. "*FlowPut: Environment-Aware Interactivity for Tangible 3D Objects*." In: *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2.1 (2018). DOI: 10.1145/3191763. URL: https://doi.org/10.1145/3191763.

[195] Michael Rohs and Georg Essl. "*Sensing-based interaction for information navigation on handheld displays*." In: *Proceedings of the 9th international Conference on Human Computer interaction with Mobile Devices and Services*. 2007, pp. 387–394.

[196] Michael Rohs and Antti Oulasvirta. "*Target Acquisition with Camera Phones When Used as Magic Lenses*." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '08. New York, NY, USA: Association for Computing Machinery, 2008, 1409–1418. ISBN: 9781605580111. DOI: 10.1145/1357054.1357275. URL: https://doi.org/10.1145/1357054.1357275.

[197] Michael Rohs, Johannes Schöning, Martin Raubal, Georg Essl, and Antonio Krüger. "*Map Navigation with Mobile Devices: Virtual versus Physical Movement with and without Visual Context*." In: *Proceedings of the 9th International Conference on Multimodal Interfaces*. ICMI '07. New York, NY, USA: Association for Computing Machinery, 2007, 146–153. ISBN: 9781595938176. DOI: 10.1145/1322192.1322219. URL: https://doi.org/10.1145/1322192.1322219.

[198] David A Rosenbaum. *Human motor control*. Academic press, 2009.

[199] Jaime Ruiz and Edward Lank. "*Analyzing Intended Use Effects in Target Acquisition*." In: *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces*. AVI '14. New York, NY, USA: ACM, 2014, pp. 145–152. ISBN: 978-1-4503-2775-6. DOI: 10.1145/2598153.2598158. URL: http://doi.acm.org/10.1145/2598153.2598158.

[200] Enrico Rukzio, Paul Holleis, and Hans Gellersen. "*Personal Projectors for Pervasive Computing*." In: *IEEE Pervasive Computing* 11.2 (2012), pp. 30–37. DOI: 10.1109/MPRV.2011.17.

[201] David E Rumelhart and Donald A Norman. "*Simulating a skilled typist: A study of skilled cognitive-motor performance*." In: *Cognitive science* 6.1 (1982), pp. 1–36.

[202] Debashis Saha, Amitava Mukherjee, Somprakash Bandyopadhyay, Debashis Saha, Amitava Mukherjee, and Somprakash Bandyopadhyay. "*Pervasive computing*." In: *Networking Infrastructure for Pervasive Computing: Enabling Technologies and Systems* (2003), pp. 1–37.

[203] Marc Satkowski, Rufat Rzayev, Eva Goebel, and Raimund Dachselt. "*ABOVE & BELOW: Investigating Ceiling and Floor for Augmented Reality Content Placement*." In: *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 2022, pp. 518–527. DOI: 10.1109/ISMAR55827.2022.00068.

[204] Jürgen Scheible, Achim Hoth, Julian Saal, and Haifeng Su. "*Displaydrone: A Flying Robot Based Interactive Display*." In: *Proceedings of the 2nd ACM International Symposium on Pervasive Displays*. PerDis '13. New York, NY, USA: Association for Computing Machinery, 2013, 49–54. ISBN: 9781450320962. DOI: 10.1145/2491568.2491580. URL: https://doi.org/10.1145/2491568.2491580.

[205] Albrecht Schmidt. "*Implicit human computer interaction through context*." In: *Personal technologies* 4 (2000), pp. 191–199.

[206] Dominik Schmidt, David Molyneaux, and Xiang Cao. "*PICOntrol: Using a Handheld Projector for Direct Control of Physical Devices through Visible Light*." In: *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*. UIST '12. New York, NY, USA: Association for Computing Machinery, 2012, 379–388. ISBN: 9781450315807. DOI: 10.1145/2380116.2380166. URL: https://doi.org/10.1145/2380116.2380166.

[207] Daniel Schneider et al. "*Accuracy Evaluation of Touch Tasks in Commodity Virtual and Augmented Reality Head-Mounted Displays*." In: *Proceedings of the 2021 ACM Symposium on Spatial User Interaction*. SUI '21. New York, NY, USA: Association for Computing Machinery, 2021. ISBN: 9781450390910. DOI: 10.1145/3485279.3485283. URL: https://doi.org/10.1145/3485279.3485283.

[208] Johannes Schöning, Michael Rohs, Sven Kratz, Markus Löchtefeld, and Antonio Krüger. "*Map Torchlight: A Mobile Augmented Reality Camera Projector Unit*." In: *CHI '09 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '09. New York, NY, USA: Association for Computing Machinery, 2009, 3841–3846. ISBN: 9781605582474. DOI: 10.1145/1520340.1520581. URL: https://doi.org/10.1145/1520340.1520581.

[209] Christian Seegelke, Charmayne ML Hughes, and Thomas Schack. "*An investigation into manual asymmetries in grasp behavior and kinematics during an object manipulation task*." In: *Experimental Brain Research* 215.1 (2011), pp. 65–75.

[210] Marcos Serrano, Anne Roudaut, and Pourang Irani. "*Investigating Text Legibility on Non-Rectangular Displays (2016)*." In: *Proceedings of CHI*. Vol. 16. 2016.

[211] Sarah Sharples, Sue Cobb, Amanda Moody, and John R Wilson. "*Virtual reality induced symptoms and effects (VRISE): Comparison of head mounted display (HMD), desktop and projection display systems*." In: *Displays* 29.2 (2008), pp. 58–69.

[212] William J. Shelstad, Dustin C. Smith, and Barbara S. Chaparro. "*Gaming on the Rift: How Virtual Reality Affects Game User Satisfaction*." In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 61.1 (2017), pp. 2072–2076. DOI: 10.1177/1541931213602001. eprint: https://doi.org/10.1177/1541931213602001. URL: https://doi.org/10.1177/1541931213602001.

[213] Yilei Shi, Haimo Zhang, Kaixing Zhao, Jiashuo Cao, Mengmeng Sun, and Suranga Nanayakkara. "*Ready, steady, touch! sensing physical contact with a finger-mounted IMU*." In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4.2 (2020), pp. 1–25.

[214] 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*. New York, NY, USA: Association for Computing Machinery, 2018, 1–11. ISBN: 9781450356206. URL: https://doi.org/10.1145/3173574.3173747.

[215] Ludwig Sidenmark, Diako Mardanbegi, Argenis Ramirez Gomez, Christopher Clarke, and Hans Gellersen. "*Bimodalgaze: Seamlessly refined pointing with gaze and filtered gestural head movement*." In: *ACM Symposium on Eye Tracking Research and Applications*. 2020, pp. 1–9.

[216] Ludwig Sidenmark, Franziska Prummer, Joshua Newn, and Hans Gellersen. "*Comparing Gaze, Head and Controller Selection of Dynamically Revealed Targets in Head-mounted Displays*." In: *IEEE Transactions on Visualization and Computer Graphics* (2023), pp. 1–11. DOI: 10.1109/TVCG.2023.3320235.

[217] Hyunyoung Song, Tovi Grossman, George Fitzmaurice, François Guimbretiere, Azam Khan, Ramtin Attar, and Gordon Kurtenbach. *"PenLight: Combining a Mobile Projector and a Digital Pen for Dynamic Visual Overlay."* In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '09. New York, NY, USA: Association for Computing Machinery, 2009, 143–152. ISBN: 9781605582467. DOI: 10.1145/1518701.1518726. URL: https://doi.org/10.1145/1518701.1518726.

[218] Hyunyoung Song, Francois Guimbretiere, Tovi Grossman, and George Fitzmaurice. *"MouseLight: Bimanual Interactions on Digital Paper Using a Pen and a Spatially-Aware Mobile Projector."* In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '10. New York, NY, USA: Association for Computing Machinery, 2010, 2451–2460. ISBN: 9781605589299. DOI: 10.1145/1753326.1753697. URL: https://doi.org/10.1145/1753326.1753697.

[219] Maximilian Speicher, Brian D. Hall, and Michael Nebeling. *"What is Mixed Reality?"* In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. New York, NY, USA: Association for Computing Machinery, 2019, 1–15. ISBN: 9781450359702. DOI: 10.1145/3290605.3300767. URL: https://doi.org/10.1145/3290605.3300767.

[220] Katta Spiel, Oliver L. Haimson, and Danielle Lottridge. *"How to Do Better with Gender on Surveys: A Guide for HCI Researchers."* In: *Interactions* 26.4 (2019), 62–65. ISSN: 1072-5520. DOI: 10.1145/3338283. URL: https://doi.org/10.1145/3338283.

[221] 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 SIGCHI Conference on Human Factors in Computing Systems*. 2014, pp. 1113–1122.

[222] Martin Spindler, Sophie Stellmach, and Raimund Dachselt. *"PaperLens: Advanced Magic Lens Interaction above the Tabletop."* In: *Proc. of the ACM Int. Conf. on Interactive Tabletops and Surfaces*. ITS '09. New York, NY, USA: ACM, 2009, 69–76. ISBN: 9781605587332. DOI: 10.1145/1731903.1731920. URL: https://doi.org/10.1145/1731903.1731920.

[223] John C Tang and Scott Minneman. *"VideoWhiteboard: video shadows to support remote collaboration."* In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. 1991, pp. 315–322.

[224] Brandon T Taylor and V Michael Bove Jr. *"Graspables: grasp-recognition as a user interface."* In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2009, pp. 917–926.

[225] Robert J. Teather and Wolfgang Stuerzlinger. *"Pointing at 3D targets in a stereo head-tracked virtual environment."* In: *2011 IEEE Symposium on 3D User Interfaces (3DUI)*. 2011, pp. 87–94. DOI: 10.1109/3DUI.2011.5759222.

[226] Martin Tomitsch, Thomas Grechenig, Andrew Vande Moere, and Sheldon Renan. *"Information Sky: Exploring the Visualization of Information on Architectural Ceilings."* In: *2008 12th International Conference Information Visualisation*. 2008, pp. 100–105. DOI: 10.1109/IV.2008.81.

[227] Eleftherios Triantafyllidis and Zhibin Li. *"The Challenges in Modeling Human Performance in 3D Space with Fitts' Law."* In: *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI EA '21. New York, NY, USA: Association for Computing Machinery, 2021. ISBN: 9781450380959. DOI: 10.1145/3411763.3443442. URL: https://doi.org/10.1145/3411763.3443442.

[228] Michael Tsang, George W. Fitzmzurice, Gordon Kurtenbach, Azam Khan, and Bill Buxton. *"Boom chameleon: simultaneous capture of 3D viewpoint, voice and gesture annotations on a spatially-aware display."* In: *ACM Trans. Graph.* 22.3 (2003), p. 698. ISSN: 0730-0301. DOI: 10.1145/882262.882329. URL: https://doi.org/10.1145/882262.882329.

[229] B. Ullmer and H. Ishii. *"Emerging frameworks for tangible user interfaces."* In: *IBM Systems Journal* 39.3.4 (2000), pp. 915–931. DOI: 10.1147/sj.393.0915.

[230] Brygg Ullmer and Hiroshi Ishii. *"The metaDESK: models and prototypes for tangible user interfaces."* In: *Proceedings of the 10th annual ACM symposium on User interface software and technology*. 1997, pp. 223–232.

[231] Dimitar Valkov, Pascal Kockwelp, Florian Daiber, and Antonio Krüger. *"Reach prediction using finger motion dynamics."* In: *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 2023, pp. 1–8.

[232] Michiel Van Veldhuizen and Xubo Yang. *"The Effect of Semi-Transparent and Interpenetrable Hands on Object Manipulation in Virtual Reality."* In: *IEEE Access* 9 (2021), pp. 17572–17583. DOI: 10.1109/ACCESS.2021.3050881.

[233] Eduardo Velloso, Jason Alexander, Andreas Bulling, and Hans Gellersen. *"Interactions Under the Desk: A Characterisation of Foot Movements for Input in a Seated Position."* In: vol. 9296. 2015, pp. 384–401. DOI: 10.1007/978-3-319-22701-6_29.

[234] 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. New York, NY, USA: Association for Computing Machinery, 2004, 137–146. ISBN: 1581139578. DOI: 10.1145/1029632.1029656. URL: https://doi.org/10.1145/1029632.1029656.

[235] Daniel Vogel and Géry Casiez. *"Hand occlusion on a multi-touch tabletop."* In: *Proceedings of the SIGCHI conference on human factors in computing systems*. 2012, pp. 2307–2316.

[236] Stephen Voida, Mark Podlaseck, Rick Kjeldsen, and Claudio Pinhanez. *"A Study on the Manipulation of 2D Objects in a Projector/Camera-Based Augmented Reality Environment."* In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '05. New York, NY, USA: Association for Computing Machinery, 2005, 611–620. ISBN: 1581139985. DOI: 10.1145/1054972.1055056. URL: https://doi.org/10.1145/1054972.1055056.

[237] Manuela Waldner, Raphael Grasset, Markus Steinberger, and Dieter Schmalstieg. *"Display-Adaptive Window Management for Irregular Surfaces."* In: *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*. ITS '11. New York, NY, USA: Association for Computing Machinery, 2011, 222–231. ISBN: 9781450308717. DOI: 10.1145/2076354.2076394. URL: https://doi.org/10.1145/2076354.2076394.

[238] Feng Wang and Xiangshi Ren. *"Empirical evaluation for finger input properties in multi-touch interaction."* In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '09. New York, NY, USA: Association for Computing Machinery, 2009, 1063–1072. ISBN: 9781605582467. DOI: 10.1145/1518701.1518864. URL: https://doi.org/10.1145/1518701.1518864.

[239] Viveka Weiley and Matt Adcock. *"Drawing in the lamposcope."* In: *Proceedings of the 9th ACM Conference on Creativity & Cognition*. 2013, pp. 382–383.

[240] Mark Weiser. *"The Computer for the 21 st Century."* In: *Scientific american* 265.3 (1991), pp. 94–105.

[241] Pierre Wellner. *"The DigitalDesk Calculator: Tangible Manipulation on a Desk Top Display."* In: *Proceedings of the 4th Annual ACM Symposium on User Interface Software and Technology*. UIST '91. New York, NY, USA: Association for Computing Machinery, 1991, 27–33. ISBN: 0897914511. DOI: 10.1145/120782.120785. URL: https://doi.org/10.1145/120782.120785.

[242] Pierre Wellner. *"Interacting with Paper on the DigitalDesk."* In: *Commun. ACM* 36.7 (1993), 87–96. ISSN: 0001-0782. DOI: 10.1145/159544.159630. URL: https://doi.org/10.1145/159544.159630.

[243] Pierre Wellner, Wendy Mackay, and Rich Gold. *"Back to the real world."* In: *Communications of the ACM* 36.7 (1993), pp. 24–26.

[244] Johann Wentzel, Fraser Anderson, George Fitzmaurice, Tovi Grossman, and Daniel Vogel. *"SwitchSpace: Understanding Context-Aware Peeking Between VR and Desktop Interfaces."* In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*. CHI '24. New York, NY, USA: Association for Computing Machinery, 2024. ISBN: 9798400703300. DOI: 10.1145/3613904.3642358. URL: https://doi.org/10.1145/3613904.3642358.

[245] Daniel Wigdor, Gerald Penn, Kathy Ryall, Alan Esenther, and Chia Shen. "*Living with a Tabletop: Analysis and Observations of Long Term Office Use of a Multi-Touch Table*." In: *Second Annual IEEE International Workshop on Horizontal Interactive Human-Computer Systems (TABLETOP'07)*. 2007, pp. 60–67. DOI: 10.1109/TABLETOP.2007.33.

[246] Ej Williams. "*Experimental Designs Balanced for the Estimation of Residual Effects of Treatments*." In: 1949. DOI: 10.1071/CH9490149.

[247] Karl D. D. Willis, Takaaki Shiratori, and Moshe Mahler. "*HideOut: Mobile Projector Interaction with Tangible Objects and Surfaces*." In: *Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction*. TEI '13. New York, NY, USA: Association for Computing Machinery, 2013, 331–338. ISBN: 9781450318983. DOI: 10.1145/2460625.2460682. URL: https://doi.org/10.1145/2460625.2460682.

[248] Karl D.D. Willis, Ivan Poupyrev, Scott E. Hudson, and Moshe Mahler. "*SideBySide: Ad-Hoc Multi-User Interaction with Handheld Projectors*." In: *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*. UIST '11. New York, NY, USA: Association for Computing Machinery, 2011, 431–440. ISBN: 9781450307161. DOI: 10.1145/2047196.2047254. URL: https://doi.org/10.1145/2047196.2047254.

[249] Karl D.D. Willis, Ivan Poupyrev, and Takaaki Shiratori. "*Motionbeam: A Metaphor for Character Interaction with Handheld Projectors*." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '11. New York, NY, USA: Association for Computing Machinery, 2011, 1031–1040. ISBN: 9781450302289. DOI: 10.1145/1978942.1979096. URL: https://doi.org/10.1145/1978942.1979096.

[250] Andrew D. Wilson and Hrvoje Benko. "*Combining Multiple Depth Cameras and Projectors for Interactions on, above and between Surfaces*." In: *Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology*. UIST '10. New York, NY, USA: Association for Computing Machinery, 2010, 273–282. ISBN: 9781450302715. DOI: 10.1145/1866029.1866073. URL: https://doi.org/10.1145/1866029.1866073.

[251] Andrew Wilson, Hrvoje Benko, Shahram Izadi, and Otmar Hilliges. "*Steerable Augmented Reality with the Beamatron*." In: *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*. UIST '12. New York, NY, USA: Association for Computing Machinery, 2012, 413–422. ISBN: 9781450315807. DOI: 10.1145/2380116.2380169. URL: https://doi.org/10.1145/2380116.2380169.

[252] Jacob O. Wobbrock, Edward Cutrell, Susumu Harada, and I. Scott MacKenzie. "*An Error Model for Pointing Based on Fitts' Law*." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '08. New York, NY, USA: Association for Computing Machinery, 2008, 1613–1622. ISBN: 9781605580111. DOI: 10.1145/1357054.1357306.

[253] Jacob O Wobbrock, Leah Findlater, Darren Gergle, and James J Higgins. "*The aligned rank transform for nonparametric factorial analyses using only anova procedures*." In: *Proceedings of the SIGCHI conference on human factors in computing systems*. 2011, pp. 143–146.

[254] Jacob O. Wobbrock and Julie A. Kientz. "*Research contributions in human-computer interaction*." In: *Interactions* 23.3 (2016), 38–44. ISSN: 1072-5520. DOI: 10.1145/2907069. URL: https://doi.org/10.1145/2907069.

[255] Dennis Wolf, Jan Gugenheimer, Marco Combosch, and Enrico Rukzio. "*Understanding the heisenberg effect of spatial interaction: A selection induced error for spatially tracked input devices*." In: *Proceedings of the 2020 CHI conference on human factors in computing systems*. 2020, pp. 1–10.

[256] Haijun Xia, Sebastian Herscher, Ken Perlin, and Daniel Wigdor. "*Spacetime: Enabling Fluid Individual and Collaborative Editing in Virtual Reality*." In: *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*. UIST '18. New York, NY, USA: Association for Computing Machinery, 2018, 853–866. ISBN: 9781450359481. DOI: 10.1145/3242587.3242597. URL: https://doi.org/10.1145/3242587.3242597.

[257] Robert Xiao, Scott Hudson, and Chris Harrison. "*WorldKit: Rapid and Easy Creation of Ad-Hoc Interactive Applications on Everyday Surfaces*." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '13. New York, NY, USA: Association for Computing Machinery, 2013, 879–888. ISBN: 9781450318990. DOI: 10.1145/2470654.2466113. URL: https://doi.org/10.1145/2470654.2466113.

[258] Robert Xiao, Scott Hudson, and Chris Harrison. "*DIRECT: Making Touch Tracking on Ordinary Surfaces Practical with Hybrid Depth-Infrared Sensing*." In: *Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces*. ISS '16. New York, NY, USA: Association for Computing Machinery, 2016, 85–94. ISBN: 9781450342483. DOI: 10.1145/2992154.2992173. URL: https://doi.org/10.1145/2992154.2992173.

[259] Robert Xiao, Scott Hudson, and Chris Harrison. "*Supporting responsive cohabitation between virtual interfaces and physical objects on everyday surfaces*." In: *Proceedings of the ACM on Human-Computer Interaction* 1.EICS (2017), pp. 1–17.

[260] Ruopeng Xu, Toshiki Sato, Shio Miyafuji, and Hideki Koike. "*OmniLantern: Design and Implementations of a Portable and Coaxial Omnidirectional Projector-Camera System*." In: *Proc. of the 2022 Int. Conf. on Advanced Visual Interfaces*. AVI 2022. New York, NY, USA: ACM, 2022. ISBN: 9781450397193. DOI: 10.1145/3531073.3531126. URL: https://doi.org/10.1145/3531073.3531126.

[261] Goshiro Yamamoto and Kosuke Sato. "*PALMbit: A Body Interface Utilizing Light Projection onto Palms*." In: *The Journal of The Institute of Image Information and Television Engineers* 61 (2007), pp. 797–804. DOI: 10.3169/itej.61.797.

[262] Xing-Dong Yang, Edward Mak, David McCallum, Pourang Irani, Xiang Cao, and Shahram Izadi. "*LensMouse: Augmenting the Mouse with an Interactive Touch Display*." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '10. New York, NY, USA: Association for Computing Machinery, 2010, 2431–2440. ISBN: 9781605589299. DOI: 10.1145/1753326.1753695. URL: https://doi.org/10.1145/1753326.1753695.

[263] Ka-Ping Yee. "*Peephole Displays: Pen Interaction on Spatially Aware Handheld Computers*." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '03. New York, NY, USA: Association for Computing Machinery, 2003, 1–8. ISBN: 1581136307. DOI: 10.1145/642611.642613. URL: https://doi.org/10.1145/642611.642613.

[264] Ye Yuan et al. "*Understanding Multi-Device Usage Patterns: Physical Device Configurations and Fragmented Workflows*." In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. CHI '22. New York, NY, USA: Association for Computing Machinery, 2022. ISBN: 9781450391573. DOI: 10.1145/3491102.3517702. URL: https://doi.org/10.1145/3491102.3517702.

[265] Shumin Zhai. "*Characterizing computer input with Fitts' law parameters—the information and non-information aspects of pointing*." In: *International Journal of Human-Computer Studies* 61.6 (2004). Fitts' law 50 years later: applications and contributions from human-computer interaction, pp. 791–809. ISSN: 1071-5819. DOI: https://doi.org/10.1016/j.ijhcs.2004.09.006. URL: https://www.sciencedirect.com/science/article/pii/S1071581904001041.

[266] Yu Zhang, Jingwei Sun, Qicheng Ding, Liuxin Zhang, Qianying Wang, Xin Geng, and Yong Rui. "*Towards Workplace Metaverse: A Human-Centered Approach for Designing and Evaluating XR Virtual Displays*." In: *International Journal of Human–Computer Interaction* 40.8 (2024), pp. 2083–2098.

[267] Jingjie Zheng and Daniel Vogel. "*Finger-aware shortcuts*." In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 2016, pp. 4274–4285.

[268] Oleg Špakov and Darius Miniotas. "*Gaze-Based Selection of Standard-Size Menu Items*." In: *Proceedings of the 7th International Conference on Multimodal Interfaces*. ICMI '05. New York, NY, USA: Association for Computing Machinery, 2005, 124–128. ISBN: 1595930280. DOI: 10.1145/1088463.1088486. URL: https://doi.org/10.1145/1088463.1088486.

APPENDICES

APPENDIX A

A.1 Study tasks and questions for Chapter 3

A.1.1 *Phase 1*

Tasks

1. In one of the applications shown on the physical display, change the icon to apple and the background color to yellow with the mouse. Move the application from the display to the table using the lamp.
2. In one of the applications shown on the physical display, change the icon to dog and the background color to purple with the mouse. Move the application from the display to the wall using the lamp.
3. In the last application shown on the physical display, change the icon to bicycle and the background color to blue with the mouse. Move the application from the display to the ceiling using the lamp.
4. Find the application with apple and move it to the wall. Do not overlay with other applications.
5. Find the application with bicycle and move the application to the table. Do not overlay with other applications.
6. Find the application with dog and move the application to the ceiling. Do not overlay with other applications.
7. Find the application with bicycle and change the left text to "High" using the lamp and keep this text after releasing the lamp.
8. Change the icon of this application from bicycle to banana with the mouse.
9. Find the application with banana and change the left text to "Low" using the lamp and keep this text after releasing the lamp.
10. Find the application with banana and change background color to yellow with the mouse.
11. Move the application with banana from the table to the phone.
12. Wear the marker on the index finger. Change the icon on the phone to car by touching the buttons next to the phone.
13. Change the background color using the slider on the phone by finger.
14. Pick up the application back from the phone to the table.
15. Move the application with car from the table back to the display.

Interview Questions

- How do you think of moving the window from the display to each surface (e.g., the wall, the ceiling, and the table)? Do you find it hard or easy?
- How do you think of moving the window from the surface to the phone?
- How do you think of the whole application flow, from the display to the surface, then to the phone, and finally back to the display using the lamp?
- How do you think of interacting with content by adjusting the lamp's height?
- Did you notice the visual change in the lamp display?
- Did you use the radar view (visual guidance) to find the object?
- How do you think of using the mouse on any surfaces under the lamp?
- What do you think of the touches on surfaces and on the phone?
- Can you think of any scenarios or applications in which this lamp system can be used?

A.1.2 *Phase 2*

Calendar

1. You want to have the calendar application visible on the surface and won't need to switch applications on the display.
2. You are now interested in today's events. Change the calendar view to a daily view.
3. Let's find the event that happens at 16 p.m. using the mouse.
4. A co-worker comes in, move the calendar application to a surface and share your schedule with them.
5. Instead of showing the entire window, move the calendar application to the clock to show events around the clock.
6. You plan to leave the seat and have an idea of your next event.
7. Move the calendar application to the phone, and you can see the events on the phone.
8. Place the phone on the table. Touch the button on the phone or on the desk to show events from the previous day on the phone.

Music Player

1. You want to play the music and have it visible on any surface.
2. Explore different interfaces for music control.
3. Change the music from "the Weekend" with the mouse.
4. A co-worker comes in and moves the application to the speaker to share this music.
5. The music can be visualised by placing the window on the ceiling. You now want to enhance the atmosphere so the co-worker can also enjoy it.

Drawing on Floorplan

1. You want to annotate a floor plan on the desk.
2. You want to see the details of the whole floor.
3. Now, you want to try a new furniture arrangement in the bedroom. Find the bedroom, and then see the details of it.
4. Draw a bed at a location that you like in the bedroom.
5. Now, you want to see the 3D rendering of the floor plan.
6. Now, you want to display the 3D design to your co-workers so they can also see it.

Interview Questions

- What do you think of these three applications?
- What changes or improvements do you think we should make to improve the lamp system or the applications?
- Suppose you have this lamp system, how will you use it (e.g. manipulating content or only displaying content)?
- Can you think of any scenarios or applications in which this lamp system can be used?
- Can you comment on your overall experience using the lamp system?

GLOSSARY

adaptability	a system's capability to modify its behavior or output based on this contextual information and changing user needs. 1
Augmented Reality (AR)	Virtual elements over-layed on the real world. 11, 12
context-awareness	a system's ability to sense and respond to an entity, which is a person, location or object that is involved in the interaction[2]. 1
Desktop computing	it refers to a computing space based on a computer, a monitor, keyboard and mouse.. 1
head-mounted displays	a display device that worn on the head. 12
Mixed Reality (MR)	Anchored virtual elements that can interact with the real world. 12
Spatial Augmented Reality (SAR)	a technology that uses projector-camera units to display virtual information on surfaces. 8, 11, 25, 104