Investigating the Ergonomics of Hybrid User Interfaces for 3D Selection in Mixed Reality

Sebastian Hubenschmid
Center for Anytime Anywhere Analytics
Aarhus University
Denmark

Paul Rinau
HCI Group
University of Konstanz
Germany

Johannes Zagermann
HCI Group
University of Konstanz
Germany

Tiare Feuchtner
HCI Group
University of Konstanz
Germany

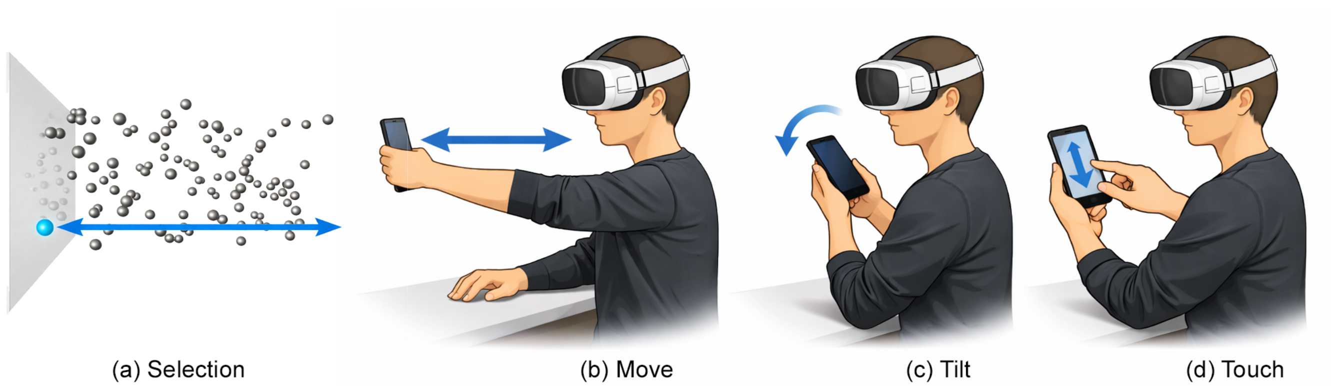

Figure 1: To support ergonomic and accurate 3D selection in a cluttered space, a movable 2D plane (a) presents a vertical "slice" of objects on the smartphone screen, enabling the target to be selected by touch. To control the plane's distance from the user, we propose three smartphone-based techniques: (b) physically *moving* the smartphone, (c) *tilting* the smartphone, and (d) panning *touch* gestures on the smartphone.

Abstract

3D selection in mixed reality environments remains challenging, as interaction techniques must carefully balance ergonomics, selection performance, and object occlusion. Recent work shows that a two-dimensional selection plane can serve as an effective intermediary, reducing 3D selection to a constrained 2D interaction space. However, existing approaches have primarily explored this concept in the context of gaze- and hand-based input, which are prone to tracking inaccuracies. In this work, we investigate how 3D selection via a 2D plane can instead be decomposed into established techniques for touch-based selection and panning on smartphones in a hybrid user interface with a head-mounted display. By leveraging the familiarity, precision, and ergonomic advantages of touch-based interaction, our approach has the potential to improve both accuracy and usability for 3D selection. To this end, we propose and discuss three techniques for controlling the movement of the 2D selection plane within 3D space, laying the groundwork for future empirical evaluation of their effectiveness.

CCS Concepts

• **Human-centered computing → Mixed / augmented reality**.

Keywords

Mixed Reality, Hybrid User Interface, 3D Selection

1 Introduction

Selecting objects is a fundamental task in mixed reality systems [8]. Especially in domains such as immersive analytics, users often need to precisely select individual items within dense, spatially distributed data. However, 3D selection through spatial interaction (e.g., controllers [15, 19] or gaze [14]) becomes increasingly difficult as the distance between users and target objects increases, as input imprecision, visual occlusion, and ergonomic constraints are amplified, thereby reducing both accuracy and usability.

To address these challenges, recent work proposed using a two-dimensional plane as an intermediary for 3D selection [14, 19]. By moving a 2D plane through the 3D scene, users can reduce the complexity of the selection task to choosing targets on a planar surface, while the third axis, defining the depth or distance of the selection, is specified by the plane's position itself. Yet, existing approaches have primarily explored this interaction using bimanual mid-air controllers [19], or free-hand input combined with gaze [14]. These techniques reportedly suffer from limitations

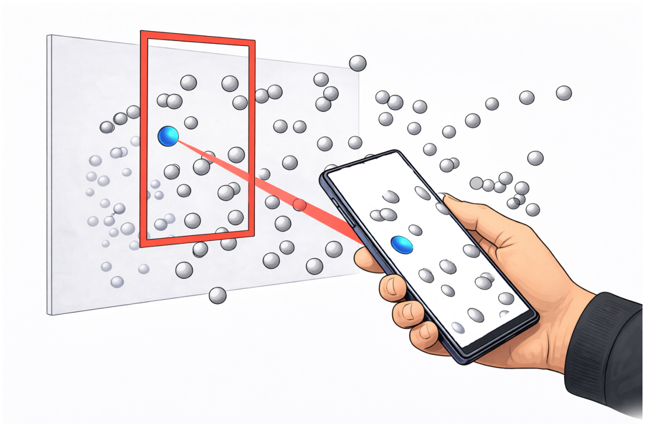

Figure 2: The smartphone displays a mirrored portion of the 2D plane, enabling object selection via touch input. The currently mirrored region is indicated by a red outline and controlled via raycasting (i.e., as shown by a red line) based on the smartphone's position and orientation.

related to fatigue, precision, and sustained ergonomic comfort, particularly during prolonged or fine-grained interaction. For example, while gaze-based interaction can offer ergonomic advantages over controller-based input, it quickly becomes error-prone in densely occluded environments [14].

In our work, we propose leveraging a hybrid user interface [5] to complement the head-mounted display with a smartphone that serves both as the controller for positioning the intermediary 2D plane and as a precise input device for selecting targets on that plane. By decomposing the task into (1) accurate 2D selection via touch-interaction and (2) linear movement of the plane within the 3D volume (see Fig. 1), we can build and extend well-established interaction techniques. We argue that this combination has the potential to provide both high selection accuracy and a familiar, ergonomically robust input modality. However, the question of how best to control the plane's distance from the user remains open, as different techniques involve distinct trade-offs in terms of precision, effort, and usability.

2 Hybrid Pointing Techniques

We employ a planar surface [19] to decompose the task of 3D selection into two subtasks: (1) defining the selection depth by moving the 2D plane through 3D space, then (2) target selection on a 2D plane. While both are well-established in prior work, the combination introduces additional challenges that require empirical investigation.

2.1 Moving the 2D Plane

To define the depth (i.e., distance from the user) of the intermediary 2D plane, we discuss three techniques based on prior work [4] for moving it linearly in 3D space: *move*, *tilt*, and *touch* (see Fig. 1). Across all techniques, the plane's motion is constrained to translation along the depth axis, as the remaining two dimensions are defined by raycasting from the user's view, which is determined by the position of the peephole (i.e., pointing the phone towards the intermediary plane, see Sec. 2.2). To avoid overlap between plane manipulation and peephole positioning in the *move* and *tilt* techniques, we implemented a *clutching* technique: plane movement is active only while a finger is touching the smartphone screen.

2.1.1 Move. While clutching, users can translate the smartphone forward and backward within their arms' reach to move the plane along the depth axis. Depending on the extent of the 3D volume, this mapping can be either 1:1 or adjusted using a control–display ratio to reduce the need for repeated clutching. Similar techniques have been explored in prior work on remote cursor control [1] and panning [4] or selection [2, 12] in three-dimensional spaces, and are also commonly found in tangible interaction paradigms [16].

2.1.2 Tilt. With this technique, users move the plane forward by tilting the smartphone down toward a more horizontal orientation, and backward by tilting it up toward a vertical orientation. To address the limited range of rotational hand motion, a control–display ratio must be applied to map phone tilting to plane movement effectively. Still, multiple clutching operations may be necessary to achieve a balance between accuracy and speed. Compared to the *move* technique, which can require substantial physical movement, tilting offers a more ergonomic alternative; however, it may occasionally interfere with raycasting, as users' tilting interferes with the smartphone's orientation. This approach builds on prior work [10, 13], which has demonstrated the effective use of tilting to control a cursor in 3D space.

2.1.3 Touch. The *touch* technique allows users to move the intermediary plane by vertically swiping the touch screen with a single finger. While this approach could offer advantages in terms of accuracy, ease of use, and low physical effort, it can also conflict with object selection on the smartphone, as the touch input is used for both tasks. This can be partially addressed by an appropriate threshold (e.g., selecting with short tap, moving with long hold and drag), but small adjustments to the plane may still result in unintended selections. Panning gestures are well-established in touch interaction and have also been successfully employed in a range of hybrid user interfaces [4, 6, 7].

2.2 Selection on 2D Plane

Prior work has explored the use of mobile devices for selecting items on distant surfaces in a variety of contexts, for example, by mirroring digital objects onto the mobile device [6] or by controlling a remote cursor [13, 17]. In particular, research on cross-device interaction has demonstrated the effectiveness of smartphones as input devices for interacting with distant 2D displays [1, 11]. Building on these approaches, we employ an intermediary 2D plane that defines a vertical slice of the target space orthogonal to the depth axis (i.e., spanning across the width and height of the target space and appearing as a screen facing the user). We use a raycasting-based technique to position a peephole over this plane, thereby projecting a portion of it onto the smartphone screen: By pointing the smartphone towards the plane, a section thereof is mirrored onto the smartphone display, enabling users to precisely select targets on the screen via touch input (see Fig. 2). This allows users to interact accurately while holding the device in a comfortable, neutral posture and to benefit from passive haptic feedback from the touchscreen.

3 Design Considerations

We discuss the design considerations of the proposed techniques, highlighting how these factors may influence their practical use and the trade-offs in usability and ergonomics.

3.1 Peephole Navigation

Although we hypothesize that using a smartphone as a hybrid user interface [5] can offer increased precision, it also introduces an additional screen that demands the user's visual attention during selection. This can lead to ergonomic challenges: For example, to position the peephole via raycasting, users may hold the smartphone as if it were a mid-air controller to keep both the device's display and the 3D space within their field of view. Such a posture can substantially increase physical demand on the arms and may lead to fatigue commonly described as the "gorilla arm syndrome" [3]. Alternatively, users may hold the smartphone in a more comfortable, low position near the hip. While this likely reduces arm strain, it can also increase neck strain and visual switching, effectively shifting ergonomic load from the arms to the neck rather than eliminating it altogether. Prior work suggests that eyes-free interaction techniques [6] could mitigate these issues. For example, providing preview cues on the intermediary plane (e.g., similar to the peephole outline) before the user's finger touches the touchscreen has been shown to reduce the need for sustained visual attention on the device, thereby alleviating some of the associated ergonomic burden [9].

Beyond ergonomics, plane-based selection may still suffer from occlusion issues in dense 3D environments, where the intermediary plane itself can be partially or fully obscured by surrounding objects. Here, transparency has been shown to be an effective technique for object selection in cluttered 3D scenes [15]; however, increased transparency can again hinder visual search by reducing object salience and making target identification more difficult.

3.2 Plane Movement

While our techniques for moving the intermediary plane build on prior work for 3D object manipulation, each introduces distinct trade-offs, particularly when combined with peephole navigation. Specifically, both the *tilt* and *move* techniques have different overlaps with the raycasting-based positioning of the peephole, while *touch* overlaps with the selection of targets. Although this can be compensated through design strategies such as clutching or thresholding, such compensations may make selection more error-prone (e.g., accidental selections) or reduce performance (e.g., as *tilt* resets the position for raycasting). Moreover, keeping conditions comparable across techniques introduces additional constraints that may limit their ecological validity. We hypothesize that, using compensation methods, the *move* technique still has the least overlap with our raycasting and selection overall, and may therefore serve as a suitable complementary approach [18]. However, this requires the most arm movement, potentially increasing physical demand and fatigue. This suggests a potential trade-off between ergonomics, performance, and precision, which warrants further investigation. We therefore plan to first examine these techniques in an initial pilot study to gain formative insights into their use, before comparing them against more established 3D selection techniques.

Finally, our current investigation is limited to linear translation of the plane along the depth axis. Prior work has demonstrated the benefits of more expressive plane manipulation in 3D space, including rotation or free movement [14]. Within our hybrid user interface, such degrees of freedom could, in principle, be mapped to the smartphone's rotation. However, doing so would likely interfere substantially with peephole navigation and might necessitate alternative selection techniques.

4 Conclusion

We propose using a smartphone as a complementary interaction device for 3D selection in head-mounted mixed reality environments. By decomposing distant 3D selection into (1) accurate 2D selection on a screen-projected intermediary plane and (2) controlled movement of that plane within the 3D space, our approach builds on well-established interaction techniques We argue that this combination, leveraging a hybrid user interface, has the potential to improve selection accuracy and overall performance compared to existing controller- and gaze-based methods, while also offering improved ergonomic comfort.

To complement raycasting-based peephole navigation on the intermediary plane, we introduce three techniques for controlling the plane's depth drawing on prior work: physically moving the smartphone, tilting it, or using touch-based panning gestures. While these techniques promise more precise and ergonomic interaction, they also introduce distinct design trade-offs that require empirical investigation. In future work, we plan to identify suitable design parameters for these techniques in a formative evaluation, followed by an empirical evaluation in a controlled user study comparing them against established 3D selection modalities [14, 19].

Acknowledgments

We thank Harald Reiterer and Ken Pfeuffer for constructive discussions on the interaction techniques. This research was funded by the *Deutsche Forschungsgemeinschaft* (DFG, German Research Foundation) – Project-ID 251654672 – TRR 161 and by the Villum Investigator grant VL-54492 by Villum Fonden. We acknowledge the use of GPT5 (OpenAI, https://chatgpt.com) for rephrasing, proofreading, and generating figures.

References

[1] Teo Babic, Harald Reiterer, and Michael Haller. 2018. Pocket6: A 6DoF Controller Based On A Simple Smartphone Application. In *Proceedings of the Symposium on Spatial User Interaction*. ACM, Berlin Germany, 2–10. doi:10.1145/3267782.3267785

[2] Lonni Besançon, Mickael Sereno, Lingyun Yu, Mehdi Ammi, and Tobias Isenberg. 2019. Hybrid Touch/Tangible Spatial 3D Data Selection. *Computer Graphics Forum* 38, 3 (2019), 553–567. doi:10.1111/cgf.13710

[3] Sebastian Boring, Marko Jurmu, and Andreas Butz. 2009. Scroll, Tilt or Move It: Using Mobile Phones to Continuously Control Pointers on Large Public Displays. In *Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7 (OZCHI '09)*. Association for Computing Machinery, New York, NY, USA, 161–168. doi:10.1145/1738826.1738853

[4] Wolfgang Büschel, Annett Mitschick, Thomas Meyer, and Raimund Dachselt. 2019. 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*. ACM Press, Taipei, Taiwan, 1–13. doi:10.1145/3338286.3340113

[5] Sebastian Hubenschmid, Marc Satkowski, Johannes Zagermann, Julián Méndez, Niklas Elmqvist, Steven Feiner, Tiare Feuchtner, Jens Emil Grønbæk, Benjamin Lee, Dieter Schmalstieg, Raimund Dachselt, and Harald Reiterer. 2026. Hybrid User Interfaces: Past, Present, and Future of Complementary Cross-Device Interaction in Mixed Reality. arXiv:2509.05491 [cs.HC] https://arxiv.org/abs/2509.05491

[6] Sebastian Hubenschmid, Johannes Zagermann, Simon Butscher, and Harald Reiterer. 2021. STREAM: Exploring the Combination of Spatially-Aware Tablets with Augmented Reality Head-Mounted Displays for Immersive Analytics. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. ACM, Yokohama Japan, 1–14. doi:10.1145/3411764.3445298

[7] Sebastian Hubenschmid, Johannes Zagermann, Daniel Leicht, Harald Reiterer, and Tiare Feuchtner. 2023. 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*. ACM, Hamburg Germany, 1–15. doi:10.1145/3544548.3581438

[8] Joseph J. LaViola, Ernst Kruijff, Ryan P. McMahan, Doug A. Bowman, and Ivan Poupyrev. 2017. *3D User Interfaces: Theory and Practice* (second edition ed.). Addison-Wesley, Boston.

[9] Jon Moeller and Andruid Kerne. 2012. ZeroTouch: An Optical Multi-Touch and Free-Air Interaction Architecture. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Austin Texas USA, 2165–2174. doi:10.1145/2207676.2208368

[10] Krzysztof Pietroszek, James R. Wallace, and Edward Lank. 2015. Tiltcasting: 3D Interaction on Large Displays Using a Mobile Device. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. ACM, Charlotte NC USA, 57–62. doi:10.1145/2807442.2807471

[11] Julian Seifert, Andreas Bayer, and Enrico Rukzio. 2013. PointerPhone: Using Mobile Phones for Direct Pointing Interactions with Remote Displays. In *Human-Computer Interaction – INTERACT 2013 (Lecture Notes in Computer Science)*, Paula Kotzé, Gary Marsden, Gitte Lindgaard, Janet Wesson, and Marco Winckler (Eds.). Springer, Berlin, Heidelberg, 18–35. doi:10.1007/978-3-642-40477-1_2

[12] Mickael Sereno, Stéphane Gosset, Lonni Besançon, and Tobias Isenberg. 2022. Hybrid Touch/Tangible Spatial Selection in Augmented Reality. *Computer Graphics Forum* 41, 3 (June 2022), 403. doi:10.1111/cgf.14550

[13] Minghui Sun, Mingming Cao, Limin Wang, and Qian Qian. 2020. PhoneCursor: Improving 3D Selection Performance With Mobile Device in AR. *IEEE Access* 8 (2020), 70616–70626. doi:10.1109/ACCESS.2020.2986037

[14] Uta Wagner, Matthias Albrecht, Andreas Asferg Jacobsen, Haopeng Wang, Hans Gellersen, and Ken Pfeuffer. 2024. Gaze, Wall, and Racket: Combining Gaze and Hand-Controlled Plane for 3D Selection in Virtual Reality. *Proceedings of the ACM on Human-Computer Interaction* 8, ISS (Oct. 2024), 189–213. doi:10.1145/3698134

[15] Jonathan Wieland, Samar Abed, Anke V. Reinschluessel, Johannes Zagermann, Harald Reiterer, and Tiare Feuchtner. 2026. Extending the Fishing Reel - Improving Multiple Object Selection in VR Using Transparency and a Resizable Pointer. In *Proceedings of the 1st International Conference on Human-Computer Interaction in the Alps (AlpCHI 2026)*. ACM, Ascona, Switzerland, 7 pages. doi:10.1145/3780045.3780046

[16] Jonathan Wieland, Maximilian Dürr, Rebecca Frisch, Melissa Michalke, Dominik Morgenstern, Harald Reiterer, and Tiare Feuchtner. 2025. Investigating the Potential of Haptic Props for 3D Object Manipulation in Handheld AR. *IEEE Transactions on Visualization and Computer Graphics* 31, 9 (2025), 6130–6146. doi:10.1109/TVCG.2024.3495021

[17] Dennis Wolf, John J. Dudley, and Per Ola Kristensson. 2018. Performance Envelopes of In-Air Direct and Smartwatch Indirect Control for Head-Mounted Augmented Reality. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, Reutlingen, 347–354. doi:10.1109/VR.2018.8448289

[18] Johannes Zagermann, Sebastian Hubenschmid, Priscilla Balestrucci, Tiare Feuchtner, Sven Mayer, Marc O. Ernst, Albrecht Schmidt, and Harald Reiterer. 2022. Complementary Interfaces for Visual Computing. *it - Information Technology* 64, 4-5 (Aug. 2022), 145–154. doi:10.1515/itit-2022-0031

[19] Futian Zhang, Keiko Katsuragawa, and Edward Lank. 2022. Conductor: Intersection-Based Bimanual Pointing in Augmented and Virtual Reality. *Proceedings of the ACM on Human-Computer Interaction* 6, ISS (Nov. 2022), 103–117. doi:10.1145/3567713