Towards Accessible XR Hand Interactions via AI-supported Interaction Proxies

Chen Liang
clumich@umich.edu
University of Michigan, Ann Arbor
Ann Arbor, MI, USA

Anhong Guo
anhong@umich.edu
University of Michigan, Ann Arbor
Ann Arbor, MI, USA

Abstract

Hand interactions are increasingly used as the primary input modality in Extended Reality (XR), but they are not always feasible due to situational impairments, motor limitations, and environmental constraints. In this position paper, I discuss an AI-supported interaction proxy approach to address these challenges. In this approach, users interact with an accessible proxy, which then carries out the required actions on the users' behalf. To illustrate this idea, I demonstrate the HandProxy system, a speech interface which, instead of relying on a fixed set of predefined speech commands, allows users to control a virtual hand, and the virtual hand serves as the interaction proxy to perform the corresponding interactions. It demonstrates the strength of using interaction proxies to support accessible hand interactions, including expressiveness, intuitiveness, and reduced modification or special API required from various XR applications.

Keywords

Interaction proxies, extended reality, accessible interactions

ACM Reference Format:
Chen Liang and Anhong Guo. 2025. Towards Accessible XR Hand Interactions via AI-supported Interaction Proxies. In *Proceedings of (CHI '25 Workshop - Everyday AR through AI-in-the-Loop)*. ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Hand tracking is increasingly used as the primary input modality on extended reality (XR) devices to interact with the virtual environment [1–3]. As a result, many applications are built specifically for hand interactions, including games, productivity apps, and 2D/3D user interfaces (UI). However, due to situational impairments [18], motor limitations [12, 19], and environmental constraints [10, 16], users may not always be able to perform the expected hand interactions, making it challenging or even impossible to effectively interact with the virtual environments. While alternatives and enhancements to hand input have been proposed in prior research [11, 13, 14, 16, 17] and adopted in mainstream XR devices such as Apple Vision Pro Voice Control [4], they are either designed to address a limited scope of specific scenarios (e.g., basic object manipulation, locomotion) or may require complex input setups, making it less practical to be deployed on existing devices for a broad range of interaction scenarios.

In this position paper, I explore AI-supported interaction proxy approaches to address these challenges, providing an accessible interface that supports expressive hand interactions while reducing the need of requiring specific accessibility support from different XR applications. I detail the overall framework of interaction proxies for XR Hand Interactions, and demonstrate an instantiation of this idea on speech interface through the control of a virtual hand as the interaction proxy.

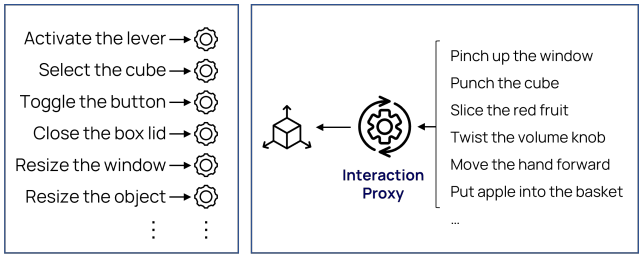

Figure 1: Existing approaches (left) where each commands need a new mapping, versus interaction proxy approach (right) that provides a unified accessible middle layer to reproduce original interactions

2 Interaction Proxies for XR Hand Interaction

Interaction proxies are the extra layer inserted between the original and the manifest interface in order to add or modify interactions without changing the app's source code [20]. Their ability to introduce new functionality with minimal modifications makes them particularly valuable for interaction remapping, such as remapping physical input with digital interactions [5–9], or hand interactions to other modalities [10, 15].

As shown in Figure 1, current approaches of improving interaction accessibility mostly focusing on creating a more accessible mapping of the original interaction to a new one. However, such approach would become cumbersome for developers, as they need to create new mappings for each supported interaction within an application. Specifically for hand interactions, many interactions are not explicitly defined but emerge as a result of other factors, such as collision and physics, which makes it impossible to simply create new mappings for all possible interactions, as interactions themselves could be loosely defined in certain modalities. Also, a "more accessible" form of input still comes with certain assumptions on user's abilities and preferences, and it is challenging to design a mapping for a wide range of users' abilities and needs.

The benefits of interaction proxies are twofold. First, by preserving the original interactions, interaction proxies can interact with the original interface with minimal modifications or support required from the target applications, which brings additional compatibility across different applications. Second, by introducing the

CHI '25 Workshop - Everyday AR through AI-in-the-Loop, Yokohama, Japan
2025. ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

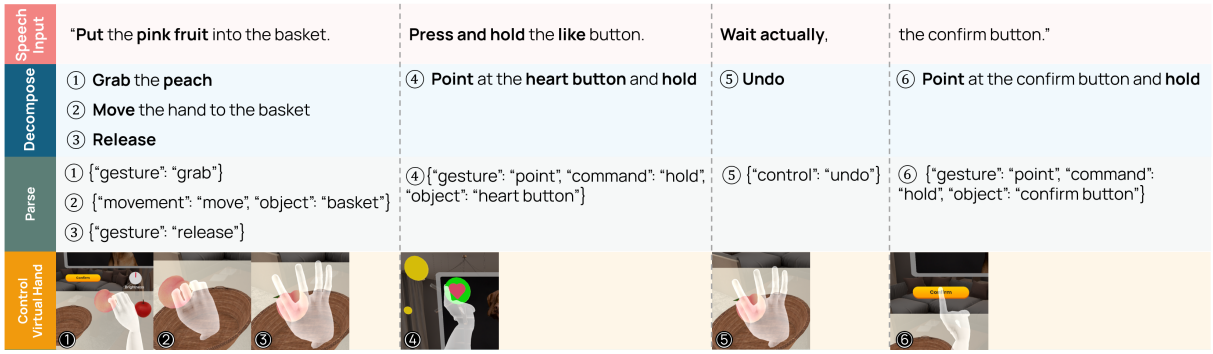

Figure 2: Example of how HandProxy would process the user's speech commands. The user continuously talks to the system, where the system concurrently recognizes the commands, decomposes it into steps, parses it into executable hand control instructions, and calculates the sequence of hand movement to control the virtual hand performing the desired interaction.

middle layer between the user and the target application, the proxy can be used as an unified interface that could be adapted to accommodate a wider range of user needs and abilities. In the context of interaction proxies in XR, several key components are critical for effective support:

(1) Proxy controls: The proxy should provide a standardized categorization of supported controls to reproduce various interactions. This ensures that interactions can be mapped to diverse user interfaces.
(2) Interaction command interpretation: The proxy should be capable of interpreting a wide range of user interaction commands, decomposing them into executable proxy controls, and performing the corresponding interactions.
(3) Contextual understanding: The proxy should analyze contextual information within the original interface, such as object states in XR environments and interaction history, to enhance command interpretation and execution.
(4) Feedback generation: The proxy should process necessary feedback from the original interface and present it in an accessible format tailored to the user's interface.

3 HandProxy: Initiating Hand Interactions Through Speech via a Virtual Proxy Hand

To illustrate this approach, I developed a speech interface enabling users to use natural language to control a *virtual hand* as an interaction proxy, where it can then simulate corresponding movements and perform necessary interactions on the user's behalf, providing a flexible alternative for users to initiate hand interactions in virtual environment. Specifically, HandProxy is built on top of a list of hand control primitives that could be used for decomposing and reproducing common hand interactions, including both detailed controls (e.g., do a pinch gesture, grab the apple) or high-level interactions (e.g., maximize the volume). The system captures users' natural speech, parses it into a list of executable commands with a Large Language Model (LLM), calculates the desired hand skeleton data, and renders it in the target system and application.

Figure 2 shows an example of how user could control HandProxy to initiate hand interactions from speech. For example, user can provide a high-level goal ("put the pink fruit into the basket"), the system decomposes it into steps, and parse each steps as executable hand controls based on the list of hand control primitives, and reproduces the hand interaction by calculating the hand pose data and control the proxy hand to perform the corresponding interaction. The user study showed that participants were able to perform various XR hand interactions through HandProxy, including mid-air gestures (e.g., swipe left), object manipulation (e.g., twist the knob), high-level interaction tasks (e.g., increase the volume), and with varying levels of complexity (e.g., one-step interactions, multi-step interactions). Using LLM, the system is able to support user's flexible speech commands and decompose user's high-level intent into executable steps. Participants were able to perform the tasks with minimum training, and were able to use their own language and prompting preferences (e.g., different word choices, descriptive commands, sentence structures), demonstrating the flexibility and intuitiveness of the system.

While this work specifically focusing on using speech interface to initiate hand interactions, given the list of hand control primitives, the proxy could also be applied in other alternative input settings, such as multimodal input setups, to accommodate a wider range of interaction needs.

4 Conclusion

In this position paper, I explored the use of AI-supported interaction proxies as an alternative approach to enhancing the accessibility of XR hand interactions. I outlined the overall structure and key components necessary to support this approach and demonstrated its application in a speech-based interface for initiating diverse hand interactions through a virtual proxy hand. This work highlights the potential of interaction proxies as a unified, compatible, and accessible interface for XR interactions, which offers greater flexibility for users with diverse needs.

References

[1] [n. d.]. Getting started with Hand and Body Tracking on Meta Quest headsets. https://www.meta.com/help/quest/articles/headsets-and-accessories/controllers-and-hand-tracking/hand-tracking/
[2] [n. d.]. HoloLens 2 gestures for authoring and navigating in Dynamics 365 Guides - Dynamics 365 Mixed Reality. https://learn.microsoft.com/en-us/dynamics365/mixed-reality/guides/authoring-gestures-hl2

[3] [n. d.]. Use gestures with Apple Vision Pro. https://support.apple.com/en-us/117741

[4] [n. d.]. Use Voice Control to interact with Apple Vision Pro. https://support.apple.com/guide/apple-vision-pro/perform-actions-with-your-voice-tan14d179ad1/visionos

[5] Mark Billinghurst, Hirokazu Kato, and Ivan Poupyrev. 2001. The MagicBook: a transitional AR interface. *Computers & Graphics* 25, 5 (2001), 745–753. doi:10.1016/S0097-8493(01)00117-0 Mixed realities - beyond conventions.

[6] Neil Chulpongsatorn, Wesley Willett, and Ryo Suzuki. 2023. HoloTouch: Interacting with Mixed Reality Visualizations Through Smartphone Proxies. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 156, 8 pages. doi:10.1145/3544549.3585738

[7] Mustafa Doga Dogan, Eric J Gonzalez, Karan Ahuja, Ruofei Du, Andrea Colaço, Johnny Lee, Mar Gonzalez-Franco, and David Kim. 2024. Augmented Object Intelligence with XR-Objects. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) *(UIST '24)*. Association for Computing Machinery, New York, NY, USA, Article 19, 15 pages. doi:10.1145/3654777.3676379

[8] Anuruddha Hettiarachchi and Daniel Wigdor. 2016. Annexing Reality: Enabling Opportunistic Use of Everyday Objects as Tangible Proxies in Augmented Reality. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1957–1967. doi:10.1145/2858036.2858134

[9] Rahul Jain, Jingyu Shi, Runlin Duan, Zhengzhe Zhu, Xun Qian, and Karthik Ramani. 2023. Ubi-TOUCH: Ubiquitous Tangible Object Utilization through Consistent Hand-object interaction in Augmented Reality. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) *(UIST '23)*. Association for Computing Machinery, New York, NY, USA, Article 12, 18 pages. doi:10.1145/3586183.3606793

[10] Mohamed Kari and Christian Holz. 2023. HandyCast: Phone-based Bimanual Input for Virtual Reality in Mobile and Space-Constrained Settings via Pose-and-Touch Transfer. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 528, 15 pages. doi:10.1145/3544548.3580677

[11] Scott Mcglashan and Tomas Axling. 1996. A Speech Interface to Virtual Environments. (11 1996).

[12] Martez Mott, John Tang, Shaun Kane, Edward Cutrell, and Meredith Ringel Morris. 2020. "I just went into it assuming that I wouldn't be able to have the full experience": Understanding the Accessibility of Virtual Reality for People with Limited Mobility. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility* (Virtual Event, Greece) *(ASSETS '20)*. Association for Computing Machinery, New York, NY, USA, Article 43, 13 pages. doi:10.1145/3373625.3416998

[13] Ivan Poupyrev, Mark Billinghurst, Suzanne Weghorst, and Tadao Ichikawa. 1996. The go-go interaction technique: non-linear mapping for direct manipulation in VR. In *Proceedings of the 9th Annual ACM Symposium on User Interface Software and Technology* (Seattle, Washington, USA) *(UIST '96)*. Association for Computing Machinery, New York, NY, USA, 79–80. doi:10.1145/237091.237102

[14] Jingze Tian, Yingna Wang, Keye Yu, Liyi Xu, Junan Xie, Franklin Mingzhe Li, Yafeng Niu, and Mingming Fan. 2024. Designing Upper-Body Gesture Interaction with and for People with Spinal Muscular Atrophy in VR. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24, Article 60)*. Association for Computing Machinery, New York, NY, USA, 1–19.

[15] Wen-Jie Tseng, Samuel Huron, Eric Lecolinet, and Jan Gugenheimer. 2023. FingerMapper: Mapping Finger Motions onto Virtual Arms to Enable Safe Virtual Reality Interaction in Confined Spaces. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23, Article 874)*. Association for Computing Machinery, New York, NY, USA, 1–14.

[16] Matt Whitlock, Ethan Harnner, Jed R. Brubaker, Shaun Kane, and Danielle Albers Szafir. 2018. Interacting with Distant Objects in Augmented Reality. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 41–48. doi:10.1109/VR.2018.8446381

[17] Adam S. Williams, Jason Garcia, and Francisco Ortega. 2020. Understanding Multimodal User Gesture and Speech Behavior for Object Manipulation in Augmented Reality Using Elicitation. *IEEE Transactions on Visualization and Computer Graphics* 26, 12 (2020), 3479–3489. doi:10.1109/TVCG.2020.3023566

[18] Jacob O. Wobbrock. 2019. Situationally-Induced Impairments and Disabilities. In *Web accessibility: A foundation for research* (2 ed.), Yeliz JYesilada and Simon Harper (Eds.). Springer, London, England, Chapter 5.

[19] Momona Yamagami, Alexandra A Portnova-Fahreeva, Junhan Kong, Jacob O. Wobbrock, and Jennifer Mankoff. 2023. How Do People with Limited Movement Personalize Upper-Body Gestures? Considerations for the Design of Personalized and Accessible Gesture Interfaces. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility* (New York, NY, USA) *(ASSETS '23)*. Association for Computing Machinery, New York, NY, USA, Article 1, 15 pages. doi:10.1145/3597638.3608430

[20] Xiaoyi Zhang, Anne Spencer Ross, Anat Caspi, James Fogarty, and Jacob O. Wobbrock. 2017. Interaction Proxies for Runtime Repair and Enhancement of Mobile Application Accessibility. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 6024–6037. doi:10.1145/3025453.3025846