Knock the Reality: Virtual Interface Registration in Mixed Reality

Weiwei Jiang*
Anhui Normal University
University of Melbourne

Difeng Yu†
University of Melbourne

Andrew Irlitti‡
University of Melbourne

Jorge Goncalves§
University of Melbourne

Vassilis Kostakos¶
University of Melbourne

Xin He‖
Anhui Normal University

Abstract

We present *Knock the Reality*, an interaction technique for virtual interface registration in mixed reality (MR). When a user knocks on a physical object, our technique identifies the object based on a knocking sound and registers a customizable virtual interface onto the object. Unlike computer vision-based methods, our approach does not require continuously processing image information. Instead, we utilize audio features which are less computationally expensive. This work presents our implementation and demonstrates an interaction scenario where a user works in MR. Overall, our method offers a simple and intuitive way to register MR interfaces.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction techniques

1 Introduction

Mixed reality (MR) headsets can provide helpful digital augmentations to the physical world. Unlike smartphones and computers, where the contents are typically enclosed within limited screen space, MR allows virtual interfaces to be placed in a 3D world. Users can easily leverage the information presented in the virtual space to complete different everyday tasks [6]. For example, when a user looks at a physical calendar, they may also have the daily schedule streamed as the virtual interface to enrich the daily tasks.

There are two essential considerations around the aforementioned MR interaction experience. First, users should be able to change their information needs on the go. That is, the virtual information should only be displayed if needed by the user. Otherwise, it must be disabled to avoid visual clutter. Second, the virtual interface should be aligned with the relevant physical object (*e.g.*, the daily schedule is close to the calendar) to provide more prompt and suitable information. To fulfill both considerations, we developed *Knock the Reality*, a technique that leverages sound and hand position information for virtual interface registration in MR.

Previous works utilizing sound show great promise in enabling new interactions [7]. For example, Laput *et al.* [5] classified sound effects for real-time user activity recognition, which can enable context-aware assistants. Chen *et al*. [3] built a system that leveraged the acoustic signals generated by sliding fingers on a table for tracking input operations. More relevant to our work, Shi *et al*. [8] utilized the knocking sound collected from a smartwatch to control Interact-of-Things (IoT) devices with passive physical objects. Gong *et al*. [4] further extended the approach to incorporate motion vibration data for more robust object identification.

*e-mail: weiwei.jiang@student.unimelb.edu.au
†e-mail: difengy@student.unimelb.edu.au
‡e-mail: andrew.irlitti@unimelb.edu.au
§e-mail: jorge.goncalves@unimelb.edu.au
¶e-mail: vassilis.kostakos@unimelb.edu.au
‖Corresponding author, e-mail: xin.he@ahnu.edu.cn

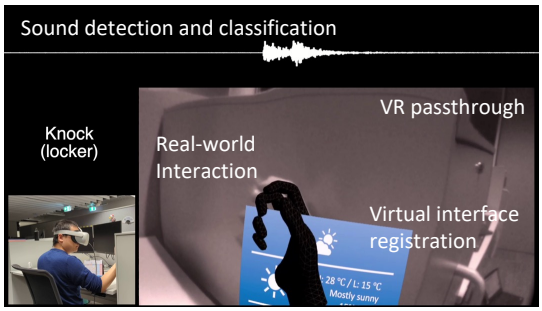

Figure 1: Overview of *Knock the Reality* in mixed reality passthrough.

Our work extends existing studies to enable new interactions in MR. In contrast to previous research, our technique enriches the user's virtual experience by using physical components. Specifically, we utilize the output from the physical world (*i.e.*, generated sound by interacting with a physical object) as the input to the virtual world through the headset's microphone. Furthermore, we combine the recognized sound information with the location of a user's hand to attach object-tailored virtual interfaces to a corresponding position in the world space. For example, when a user knocks a calendar in the physical space, a timetable will be displayed in the virtual space near the calendar (based on the hand's position when the knock was detected) to provide helpful information to the user. In addition, we highlight that our method does not rely on real-time object recognition based on computer vision, which can be computationally expensive. Our technique provides a simple and intuitive method to register MR interfaces.

In this work, we present an initial implementation of *Knock the Reality* and illustrate example applications to demonstrate its use cases in a scenario where a user can work seamlessly in both virtual and physical workspaces in MR.

2 Knock the Reality

The overview of our method is shown in Figure 1. The system consists of an MR headset and a sound detection model using machine learning. The user is in the MR passthrough mode and can see the surrounding environment. The user can register a virtual interface to the object in the MR space by producing a knocking sound on a real-world object. In this section, we show the system workflow and implementation of *Knock the Reality*.

2.1 System Workflow

The system workflow includes three steps as follows

1. Knock – The user knocks a real-world object twice to generate a knocking sound.
2. Detect – The headset's microphone receives the knocking sound that is subsequently detected by the classifier. The classifier also identifies the object that is being knocked.
3. Register – To place the virtual interface in relation to the knocked object, the interface is registered at the hand position that is captured by the built-in camera of the headset.

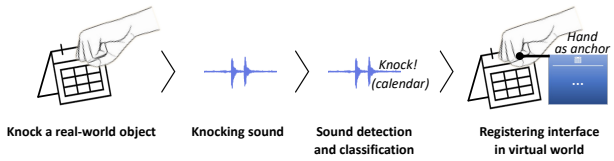

Figure 2: Workflow of *Knock the Reality*.

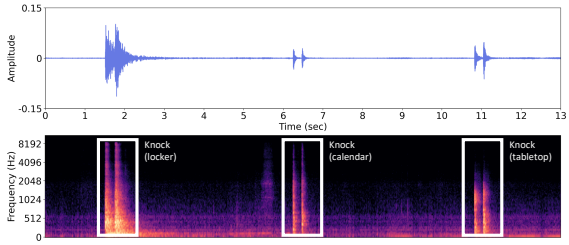

Figure 3: Illustration of sound detection for knock events.

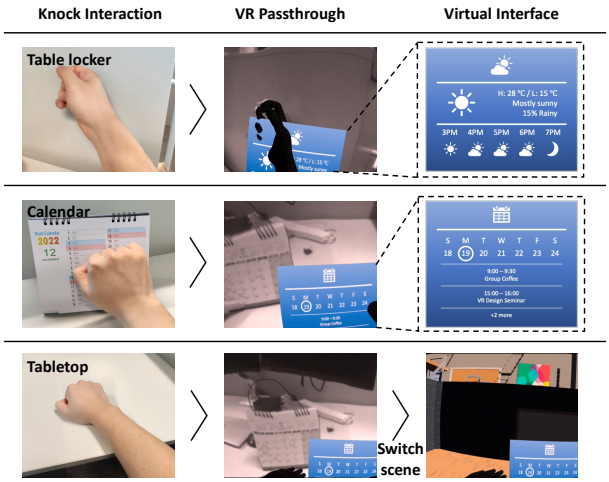

Figure 4: Examples of knock interactions in MR passthrough.

2.2 Implementation

To detect and identify the knocking sound, we trained a 2-layer convolutional neural network (CNN) model for classifying the audio inputs from the headset's microphone [1]. The classifier utilizes a sliding window with a 1-second width and a 0.8-second stride size to segment the audio stream. For demonstration, we include four sound classes – silence, knock (locker), knock (calendar), and knock (tabletop) – for different everyday objects. The illustration for sound detection is shown in Figure 3. The sound segments are converted to spectrograms as inputs to the CNN model. A registration event is triggered when a non-silence class is detected. Our implementation is based on Unity and Meta Quest 2.

3 Example Applications

To demonstrate a proof-of-concept use case, we show three examples utilizing the knock interactions in a mixed reality scenario, as illustrated in Figure 4. In particular, the user is in a workspace in both physical world and virtual world while sitting in front of a desk. When needed, the user can quickly switch to MR and bring a surrounding physical object into virtual world. For example, we consider the following use cases:

- Time and weather – The user wants to know current time and the weather, and knocks on the locker that is by the table. A virtual information panel showing hourly weather is then registered by the real-world locker.
- Calendar – The user wants to check the incoming events, and knocks a physical calendar on the table. The corresponding digital calendar with the agenda is then registered by the real-world calendar.
- Switch scenes – The user wants to switch virtual scenes or virtual environments, and knocks the tabletop. The scene is then switched to a virtual office with the virtual table registered at the user's front.

Overall, we envision that our technique can be used in the scenarios to mix the virtual and physical world with interactive digital interfaces such as information panels (*e.g.*, weather, online calendar, device status) and digital twins (*e.g.*, a registered virtual-physical furniture). We highlight that our technique can be used for bridging the physical world and the digital world in an interactive way.

4 Conclusion and Future Work

Our method provides a simple and intuitive way to register MR interfaces through knock interactions. Our proof-of-concept scenario demonstrates the use cases of our approach in an MR workspace. In future work, we aim to further expand the design space of sound-based interactions such as everyday objects with different shapes and materials. Other than everyday objects, we can incorporate *Knock the Reality* with 3D prints that can generate pre-defined sounds with user interaction [2] that can be considered as shortcuts for enhancing interactivity in mixed-reality.

Acknowledgments

This work was supported in part by the NSFC under grants No. 62072004.

References

[1] O. Abdel-Hamid, A.-r. Mohamed, H. Jiang, L. Deng, G. Penn, and D. Yu. Convolutional neural networks for speech recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 22(10):1533–1545, 2014. doi: 10.1109/TASLP.2014.2339736

[2] R. Ballagas, S. Ghosh, and J. Landay. The design space of 3d printable interactivity. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(2), jul 2018. doi: 10.1145/3214264

[3] M. Chen, P. Yang, J. Xiong, M. Zhang, Y. Lee, C. Xiang, and C. Tian. Your table can be an input panel: Acoustic-based device-free interaction recognition. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 3(1), mar 2019. doi: 10.1145/3314390

[4] T. Gong, H. Cho, B. Lee, and S.-J. Lee. Knocker: Vibroacoustic-based object recognition with smartphones. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 3(3), sep 2019. doi: 10.1145/3351240

[5] G. Laput, K. Ahuja, M. Goel, and C. Harrison. Ubicoustics: Plug-and-play acoustic activity recognition. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, UIST '18, p. 213–224. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3242587.3242609

[6] F. Lu and Y. Xu. Exploring spatial ui transition mechanisms with head-worn augmented reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3491102.3517723

[7] G. Luo, P. Yang, M. Chen, and P. Li. Hci on the table: robust gesture recognition using acoustic sensing in your hand. *IEEE Access*, 8:31481–31498, 2020. doi: 10.1109/ACCESS.2020.2973305

[8] L. Shi, M. Ashoori, Y. Zhang, and S. Azenkot. Knock knock, what's there: Converting passive objects into customizable smart controllers. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '18. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3229434.3229453