Ring System for Improving Keyboard Input and Clicking in Virtual Reality

Yichen Zhu
ychenzhu@umich.edu
University of Michigan

Jiahang Mao
bradleyx@umich.edu
University of Michigan

Andi Xu
andixu@umich.edu
University of Michigan

Jui Po Hung
rhung@umich.edu
University of Michigan

Yaxuan Li
yaxuanli@umich.edu
University of Michigan

ABSTRACT

The rapid development of virtual and augmented reality (VR/AR) technologies has been held back by limitations in interaction methods, particularly in keyboard input and clicking. In this research, we contribute a novel solution to enhance VR interactions by integrating minimal yet effective hardware components, such as a Voice PickUp Bone Sensor (VPU) microphone, a button, and a scrolling wheel, with the Meta Quest 2 finger tracking system. Our innovative approach, which emphasizes the use of the VPU microphone to capture Surface Acoustic Waves, enables accurate detection of finger contact with physical surfaces, significantly improving human-computer interaction. The added hardware components provide more accurate, versatile, and user-friendly functions, such as clicking and typing. We thoroughly evaluated the performance of our system through a comprehensive user study, combining both quantitative and qualitative methods. Our results showcase the potential for our significant contribution to improve VR/AR interaction, paving the way for future advancements in motion-tracking technology. Future work may include the addition of an inertial measurement unit (IMU) and machine learning techniques for more complex gesture recognition.

KEYWORDS

Virtual Reality, Voice PickUp Bone Sensor (VPU) microphone, Gesture recognition, Virtual keyboard

1 INTRODUCTION

The growth of virtual and augmented reality technologies has been slowed by the lack of intuitive and efficient interaction methods. For example, the Meta Quest system, which uses a purely computer vision (CV) approach for hand gesture recognition, has difficulty providing high accuracy during virtual keyboard typing. Our project aims to improve user experience by adding minimal additional hardware to increase the reliability of traditional CV systems and enhance fundamental VR interactions, like typing, making them more accurate, versatile, and user-friendly.

Current VR controllers and hand-tracking solutions have several limitations, including low accuracy, lack of haptic feedback, finger occlusion, and limited dexterity. To tackle these issues, we incorporated a Voice PickUp Bone Sensor (VPU) microphone and a button into the Meta Quest 2 finger tracking system. This combination allows for the use of a CV-based method while confirming finger-tapping actions with sensory data. The addition of these minimal hardware components enables more versatile and accurate functions, such as clicking and typing.

This project has the potential to make a significant impact on the virtual and augmented reality field by providing a more intuitive and efficient way to interact with these technologies. By addressing the limitations of current VR controllers and hand-tracking solutions, we hope to improve the user experience of virtual keyboard typing and pave the way for future advancements in motion-tracking technology. In the long run, we believe that more advanced technology could replace CV-only products, leading to a new era of highly accurate VR systems.

2 RELATED WORK

The development of intuitive and efficient methods for interacting with virtual and augmented reality has been an active research area in recent years. Researchers have proposed various approaches to overcome the limitations of current VR controllers and hand-tracking solutions, including those related to accuracy, haptic feedback, and finger dexterity.

Placidi [2] proposed the virtual glove (VG), a system employing multiple sensors surrounding the hand. Such sensors are stationary and are situated on the vertices of an imaginary cube. This can achieve more accurate and robust hand tracking, as at least two cameras can acquire 3D information due to the redundancy of sensors. This system can overcome the issue of finger occlusion, which can cause a loss of tracking data. Unfortunately, this system is unwieldy and not portable, rendering it unfeasible for the general public. It is highly unlikely that the regular user would enjoy being bound by the system, unable to feasibly bring it along.

In addition, Novacek Et al. [1] combined novel algorithms with existing technologies. The authors utilized the Leap Motion sensor to detect gestures and processed incoming data using their own hand-tracking algorithms. Specifically, they employed multiple sensors and built a custom MultiLeap library to connect to all the sensors. The algorithm is able to distinguish the most reliable, precise sensor and gives it the most weight in hand-tracking, providing a high-confidence result.

To add to the sophistication, Usakli Et al. [3] presented a way to communicate with computers using only eye movements. The authors applied an electrooculogram (EOG) to communicate with a microcontroller without speaking and hand movements, a futuristic take on virtual reality. The system processes signals from the EOG and utilizes the nearest neighborhood algorithm to classify them.

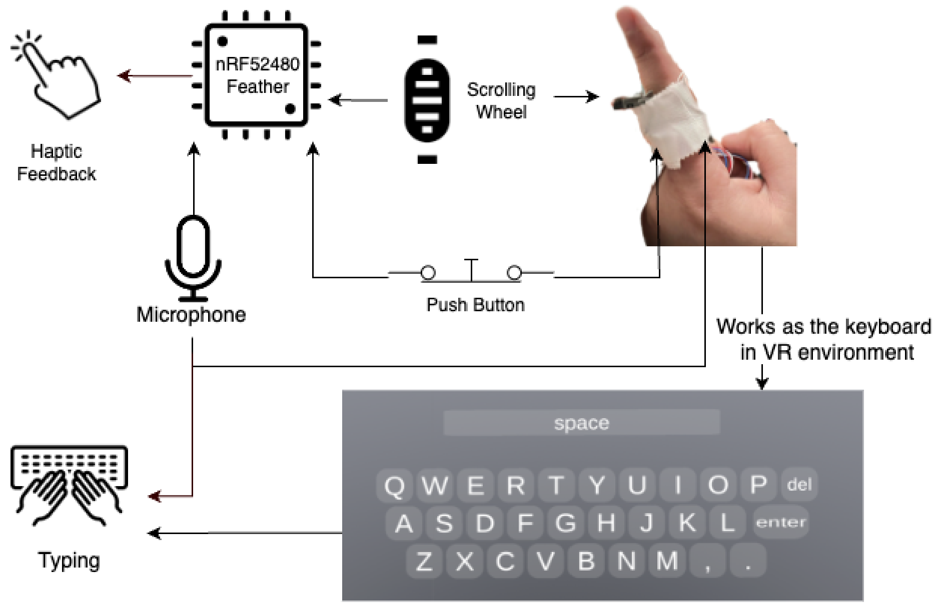

Figure 1: System Block Diagram

With a plethora of novel technologies/applications in the VR field, we aim to develop an easy-to-use, relatively simple system that provides high accuracy at the same time. We employ a widely-available CV system using the Meta Quest 2 and a custom-designed, portable piece of hardware. We believe that our system will provide a great balance of accuracy and portability.

3 SYSTEM OVERVIEW

Owing to the constraints associated with camera-based systems and hand gesture recognition, text input in virtual reality (VR) poses a significant challenge. For instance, camera systems are unable to precisely discern the moment a user makes contact with a surface, or when they are merely hovering above it at a minimal distance. To augment the user experience, we have implemented a voice processing unit (VPU) microphone on the finger, enabling the accurate detection of the instant the finger comes into contact with a physical plane. In this section, we described the detailed design of hardware and software of our system.

3.1 Hardware

In this section, we present the detailed design and implementation of the hardware. We employed the NrF524480 microprocessor to serve as the central controller for the Voice PickUp Bone Sensor (VPU), physical button, and scrolling wheel. The microprocessor's ability to perform PDM to PCM wave transformation makes it a versatile choice for a wide range of applications. The VPU captures Surface Acoustic Waves, ensuring clear communication in noisy environments. The physical button provides a reliable mechanism for interaction, while the scrolling wheel offers intuitive navigation and zooming features. Together, these hardware components are used in innovative ways to enhance audio processing and improve human-computer interaction.

A case was 3D printed to protect and enhance the appearance of the hardware. Velcro tape was used to bond the NrF52480 on the arm, while cloth tape attached the VPU microphone, physical button, and scroll wheel to the finger, establishing a secure and durable connection.

3.1.1 NrF52480 Feather: The decision to utilize the NrF524480 microprocessor was primarily based on its ability to perform Pulse-Density Modulation (PDM) to Pulse-Code Modulation (PCM) wave transformation, which is essential for high-fidelity audio processing. It is equipped with specialized hardware that enables it to process PDM signals and convert them into PCM signals. In this project, an external VPU is being used to generate PDM signals, which are then transmitted to the NrF524480 microprocessor for further processing. The compatibility of the microprocessor with the Arduino framework provides a user-friendly interface for programming and rapid prototyping, making it an ideal choice for this application.

3.1.2 Microphone: We used a Voice PickUp Bone Sensor (VPU) as our microphone. Originally developed for earbuds, the VPU is designed to capture waves that travel from the vocal cords to the inner ear tissues, ensuring clear communication in noisy environments. The sensors are hermetically sealed, allowing them to reject airborne sound waves and only capture Surface Acoustic Waves through direct contact. These sensors are fabricated using a MEMS process, making them comparable in bandwidth and size to traditional MEMS microphones found in commercial off-the-shelf devices. We are able to track finger tapping on any surface with high precision, allowing us to create a virtual keyboard and accurately detect which keys are being typed. Additionally, the microphone can detect taps on our ring, allowing for basic gestures like tapping and double tapping.

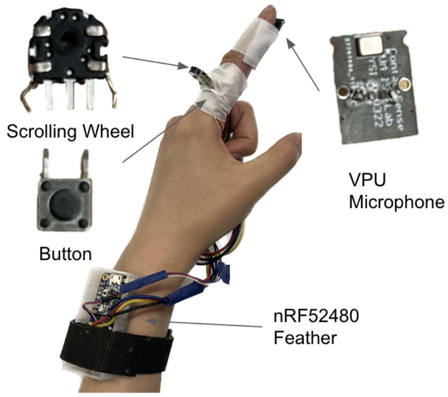

Figure 2: Processor Hardware

Table 1: Camera and VPU sensor fusion algorithm

Camera detection	VPU detection	Fusion decision
pressed	not pressed	do not trigger
not pressed	pressed	Identify the lowest finger position on the Z-axis and the nearest key to the fingertip; Increase the detection threshold for virtual plane interaction to effectively trigger button presses
pressed	pressed	trigger button presses

3.1.3 Physical Button: The utilization of a physical button has been posited as a more reliable and precise mechanism for interaction in contrast to pinch gesture recognition through a camera. As a result, an empirical examination has been undertaken to assess the speed and accuracy of both camera-based and physical button-based modalities within a virtual reality context.

3.1.4 Scrolling Wheel. The scrolling wheel was designated as a tool to scroll through pages and zoom in or out, providing a familiar and intuitive experience for users. Since traditional computer mice typically feature a scrolling wheel, users find it helpful and intuitive to have a similar feature when using VR.

3.2 Software

To optimally evaluate the performance disparity between the presence and absence of the microphone, we designed a virtual reality keyboard application as a comparative tool. By utilizing our microphone hardware, users can superimpose a virtual keyboard onto a physical surface, thereby facilitating tangible haptic feedback upon tapping the table, as opposed to aimlessly striking the air. This approach allows for a comprehensive assessment of the performance differences attributable to the hardware implementation.

In the absence of the VPU microphone, finger tapping or poking of a key button is detected solely through cameras on the VR headset. Each key button on the keyboard is associated with a virtual plane. A button is activated when the camera detects the finger moving beneath the virtual plane along the Z-axis. However, we observed that when the finger rapidly oscillates within a relatively small distance, the camera struggles to accurately determine whether a button has been triggered or not. By incorporating the VPU microphone onto the finger, we combine data from both the microphone and the camera. The algorithm employed for fusing the data from these two sensors is presented in Table 1. Given the VPU's ability to precisely detect the moment a finger makes contact with a physical plane, we assign a higher priority to the VPU sensor in our fusion approach.

As depicted in Figure 1, the user interface comprises a keyboard and a virtual display that presents the text input. Furthermore, we have integrated ChatGPT into our application, allowing users to switch to chat mode and engage in question-asking interactions with ChatGPT. Additionally, we incorporated a passthrough feature in the VR headset, facilitating the utilization of a virtual keyboard in an augmented reality environment. This functionality permits users to reposition and calibrate the keyboard using a controller, with the capacity to manipulate its orientation along the x, y, and z axes to achieve optimal placement.

4 USER STUDY

4.1 Study Aim

In an effort to increase the accuracy, speed, reliability, and user experience in a virtual setting, we work on developing our keyboard with several hardware improvements.

The goal of the user study is to confirm that including hardware leads to better gesture position tracking compared to a system that relies only on computer vision. In our study, we measured both the time and accuracy of the system to evaluate the overall user experience, taking into account the improvements made to gesture tracking.

4.2 Study Design

The experimental design of our study combines both quantitative and qualitative methods to assess the effectiveness of the developed keyboard system. For the quantitative aspect, we employed two distinct methods. Firstly, we utilized the VPU microphone and a purely virtual keyboard, requiring participants to type the same word 20 times. Subsequently, we measured the time taken and accuracy achieved for this task. Secondly, we assessed our button functionality by instructing users to interact with physical buttons and virtual pinch gestures, in order to click on 9 dots of varying sizes and positions. By recording the time and accuracy for this task, we were able to evaluate the effectiveness of our system in clicking operations. We analyzed both qualitative and quantitative data to reach a conclusion about the potential superiority of our system over the existing system. The ultimate goal of the study is to achieve positive results across all measurements, thus demonstrating the effectiveness and user satisfaction with our keyboard system.

For the qualitative aspect, we gathered user feedback through comprehensive surveys that included questions comparing our new keyboard to the existing keyboard in the Oculus environment and identifying the most significant challenges faced while using the keyboard.

The study population consists of in total of eight students (5 females and 3 males) in the Electrical Engineering and Computer Science (EECS) department who are familiar with the virtual reality environment. This ensures that participants can start using the keyboard without facing significant barriers.

The study procedure consists of the following steps:

- Participants put on the ring and the Meta Quest headset.
- The typing study for the keyboard involved participants typing the same word 20 times. This was done using the a purely virtual keyboard first, followed by a VPU microphone, with the time taken and accuracy of each attempt recorded.
- The button clicking study required participants to click on 20 dots of varying sizes and positions. The clicks were executed using two different methods: first virtual pinch gestures, which solely relied on the camera on the VR headset, and then physical buttons from our system. The study recorded the time and accuracy of each attempt for both methods.
- A survey was administered to gather participants' feedback on their experience using the controller. This feedback aims to understand our system's ease of use, comfort, and overall satisfaction with the controller.

By following this comprehensive study design, we aim to thoroughly evaluate the effectiveness of our keyboard system in a virtual environment.

4.3 User Study Results

4.3.1 Keyboard Typing User Study Results. In the following paragraph, we presented the average typing speed in words per minute (WPM) for all eight participants in our study on keyboard input, as described in Table 2. The data was collected from each participant's use of both a virtual keyboard only and a virtual keyboard with a VPU, and provides insights into the potential differences in typing performance between the two input methods.

Moreover, for the clicking study, we used Fitt's Law to compare using simply camera pinch detection with triggered by physical button. We calculated the index of difficulty (ID) using the formula:

$$ID = log_2 \frac{2D}{W} \tag{1}$$

where D is the distance from starting point to the target in each trial, and W is the width of the target. Then we calculated the regression coefficients a and b using

$$MT = a + b \times ID \tag{2}$$

where MT is the movement time in each trial. In the end, we calculated the coefficients of camera system and physical button system as shown in Table 3. A lower value of a represents a lower constant time and a lower value of b indicates a more efficient performance of the method as the index of difficulty increase. Therefore, we concluded that using a physical button could improve the accuracy and efficiency of pinching gesture detection.

Table 2: User study results for keyboard typing challenge

	Raw wpm	Freebie errors	Penalty	WPM w/ penalty
camera only	10.005	4.316	5.333	4.971
with VPU	18.710	4.316	0	19.894

Our survey of users' experiences with our VR keyboard system showed that, when using the ring system with the added microphone, users rated it higher than when using a purely computer vision based approach. Most users rated the system a 4 out of 5, with one user giving it a 3/5 rating. Users praised the system for its accuracy and futuristic feel, but reported some difficulty with certain keys. Overall, our survey showed positive user experiences and provided valuable feedback to refine and improve the design for future iterations.

Here are some user reviews:

- "*Its a good mix of traditional keyboard where you can feel the keypresses but it has the flexibility and reconfigurability of virtual reality.*"
- "*Very interesting work! I can see it useful once it gets perfected!*"
- "*It was an unexpectedly pleasant typing experience. Love the adjustment options.*"

There are certainly places where our system could improve. Users explained that some keys are a little hard to press and that our system should support using all 10 fingers simultaneously. In addition, our keyboard is slower and less accurate than the traditional physical keyboard and another user complained that the space bar is above the keyboard.

5 DISCUSSION AND FUTURE WORK

In this section, we discussed practical challenges and opportunities for applying our system. One of the most significant challenges we encountered was the instability of the Oculus camera system, which heavily relies on camera tracking to detect the user's hand movements. This instability made it impossible for us to rely purely on hardware to track the hand, and we were forced to use camera tracking in conjunction with the hardware components. However, the camera system's view of the hand could be obstructed by the

Table 3: User study results for clicking testing

	a	b
physical button	-0.15	3.31
camera only	-0.21	3.40

ring structure used to mount the VPU and other hardware components. This obstruction interfered with the camera's ability to accurately track the hand movements and posed a significant obstacle to achieving reliable results.

The ring structure itself added an extra layer of complexity to the tracking process, as its size, shape, and position needed to be taken into account during calibration and tracking. This challenge highlights the importance of considering the physical structure of the hardware components and their potential impact on the tracking process when developing solutions for virtual reality environments. In addition, the camera system's impact on calibration was another challenge we faced. The camera system's view of the hand could be affected by changes in lighting, camera position, or other external factors, leading to slight offset differences that needed to be adjusted every time the VR system was turned on. This challenge underscores the need for robust calibration methods that can account for potential external factors and ensure consistent and accurate tracking results.

One area of future improvement for our solution is the addition of an inertial measurement unit (IMU) to improve position and motion tracking. By incorporating an IMU, we could potentially enhance the accuracy and reliability of our hardware implementation, enabling more precise gesture recognition based on the VPU data collection. In addition, we could potentially eliminate the need for physical buttons and scrolling wheels to be connected to the microcontroller, as these actions could be picked up by the VPU alone. This would result in a more streamlined and intuitive user experience in VR. These potential improvements demonstrate the potential for continued development and innovation in motion tracking technology, and may pave the way for even more immersive and interactive virtual environments in the future. Another area of potential improvement for our solution is the utilization of machine learning techniques to enable more complex gesture recognition based on the VPU's ability to collect sound. With the ability to detect the instant the finger comes into contact with a physical plane, the VPU could potentially be used to recognize more complex actions such as dragging on a table or clicking on different surfaces. By implementing machine learning algorithms, we could improve the accuracy and versatility of the VPU in recognizing a wider range of gestures and actions. This could lead to even more intuitive and immersive interactions in virtual reality environments, enabling users to perform actions that are currently difficult or impossible to achieve with existing hardware and software solutions. These potential improvements demonstrate the power and potential of combining advanced hardware implementations with cutting-edge machine learning techniques, and may pave the way for even more innovative and interactive virtual environments in the future.

6 CONCLUSION

In our research, we aimed to improve the functionalities of keyboard input and clicking in virtual reality (VR) environments. To achieve this goal, we developed a solution that incorporated a Meta Quest 2 finger tracking system with a Voice PickUp Bone Sensor (VPU) microphone mounted onto a ring, as well as a button and a scrolling wheel. These additions allowed for more versatile and accurate functions, enabling users to perform actions such as clicking and scrolling with ease. By utilizing this hardware implementation, we were able to superimpose a virtual keyboard onto a physical surface, providing tangible haptic feedback upon tapping the table, and improving the user experience in VR. Our findings may contribute to the development of more intuitive and immersive technologies in motion tracking.

REFERENCES

[1] Tomas Novacek, Christian Marty, and Marcel Jirina. 2021. Project MultiLeap: Fusing Data from Multiple Leap Motion Sensors. In *2021 IEEE 7th International Conference on Virtual Reality (ICVR)*. 19–25. https://doi.org/10.1109/ICVR51878.2021.9483819

[2] Giuseppe Placidi. 2007. A smart virtual glove for the hand telerehabilitation. *Computers in biology and medicine* 37 (09 2007), 1100–7. https://doi.org/10.1016/j.compbiomed.2006.09.011

[3] Ali Bulent Usakli and Serkan Gurkan. 2010. Design of a Novel Efficient Human–Computer Interface: An Electrooculagram Based Virtual Keyboard. *IEEE Transactions on Instrumentation and Measurement* 59, 8 (2010), 2099–2108. https://doi.org/10.1109/TIM.2009.2030923