Inclusive Voice Control Interface in Virtual Reality

ANTHONY JUVERA*, Colorado State University, USA

Voice-based interfaces offers promising alternatives to handheld controllers in virtual reality (VR), especially for users who have limited motor function. In this study, I explored a hands-free VR navigation system designed in Unity for the Oculus Quest 3. The system integrates Wit.ai, a cloud-based natural language processing platform, to interpret spoken navigation commands, enabling step-based movement through a virtual maze. Users orient direction by turning their head past a predefined angular threshold, triggering reorientation. This pilot study used a small, diverse sample that tested the usability of this system with a series of maze completion tasks. Performance was evaluated using task completion time, command recognition accuracy, and post-task user feedback. (add results)

Additional Key Words and Phrases: Virtual Reality, Voice Control

ACM Reference Format:

Anthony Juvera. 2025. Inclusive Voice Control Interface in Virtual Reality. 1, 1 (April 2025), 6 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Virtual reality (VR) has become a powerful tool for immersive interaction across gaming, education, and healthcare. However, the standard method of interaction, being the use of handheld controllers, creates barriers for individuals with limited hand mobility. Traditional interfaces that depend on manual input are not well suited for users who cannot effectively manipulate these devices, thereby excluding potential users from fully engaging with VR experiences.

In order to bridge this gap, recent studies have explored alternative input modalities but many have required complex machine learning pipelines or are not optimized for standalone devices like the Oculus Quest 3. This work proposes a simplified voice-based system that integrates Wit.ai and head-tracking in order to navigate in virtual environments. Using Wit.ai, a cloud-based natural language processing platform, the system interprets spoken navigation commands, such as "forward 5 steps" to navigate through a maze-like virtual environment, while head tracking is used to make turns.

By facilitating hands-free operation, the system aims not only to improve usability but also to empower users with motor impairments to experience VR in a more natural and inclusive way. Ultimately, this research seeks to contribute to a more accessible design ensuring that immersive technologies are usable by all users regardless of their physical capabilities Prior work has not widely explored voice commands with distance modifiers or combined them with head-tracking for full navigation. This project fills that gap by building and testing a prototype that uses spoken step commands and head orientation for hands-free VR movement.

*Research conducted for an undergrad course.

Author's Contact Information: Anthony Juvera, antjuve@colostate.edu, Colorado State University, Westminster, Colorado, USA.

Manuscript submitted to ACM

2 Related Work

Since the early development of VR, natural language interaction has long been at the forefront of proposed interfaces. Though, historically underutilized compared to gaze or gesture controls. Voice input has often only been included when necessary, rather than being a standard interaction method or systematically evaluated on performance and user preference across all VR applications [1]. Early systems typically only supported simple spoken commands or short phrases combined with pointing gestures, limiting their role to supplementary interaction method.

More recent work has demonstrated the effectiveness of purely speech-driven controls for scenarios where the subjects hands are busy, like in a sterile medical setting, where voice becomes the primary means of input [5]. These studies reinforce the idea that voice interaction can be both functional and intuitive, especially when traditional input methods are impracticable. Building on these insights, the system presented in this study uses voice as the primary input modality, replacing handheld controllers entirely. Exploring how users issue natural spoken commands like "forward five steps" to control locomotion.

2.1 Usability

Evaluations of voice input in VR suggest this it has the potential to add to the user's experience [4] and serve as an effective interaction method. When voice control navigation, while in a wheelchair within virtual environments, is compared to traditional input modalities, it was revealed that participants travel longer distances but also had a higher collision rate [4]. However, combining voice control with autonomous systems enables a higher degree of interaction and control, which positively impact quality of life, efficiency, and safety [4]. Similarly, participants who were assessed on their perception of voice control across orientation, customization, and analysis tasks, also rated usability high and provided positive feedback on factors such as sickness, comfort, and presence, with an overall accuracy of 85 percent using voice input alone [5].

These outcomes support the conclusion that voice is a feasible input modality for VR environments. This study draws on those finding to evaluate a voice-controlled navigation system that relies solely on speech for directional movement, enhanced through natural phrasing and explicit step counts.

2.2 Accessibility

Despite the rapid evolutions of VR technology, accessibility remains an afterthought in most virtual environments. With over one hundred locomotion techniques available, few research efforts have explored their benefits for users with mobility impairments [7]. A 2024 review of 330 popular VR applications reported a general lack of accessibility features in current VR content [3]. Many did not even include options like audio cues or haptic feedback to convey information that would help user with limited vision or mobility. Alternative input methods are still rare in most commercial VR titles [3].

This trend has not gone unnoticed, though, accessibility experts have begun publishing guidelines geared towards immersion. For instance, the University of South Carolina's Center for Teaching Excellence acknowledges the accessibility challenges that users face and have set guidelines that are designed to explain the accessibility issues with virtual environments and provide practical solutions [6]. Similarly, industry articles emphasize providing choices like seated play modes, adjustable movement speeds, and customization button mappings to accommodate those with limited motor function [2].

Motivated by the inclusive guidelines from experts, my system emphasizes flexibility, low effort input and choice. Instead of relying on a joystick for manipulation, users speak natural commands that are parsed by Wit.ai and translated into movement, making VR navigation more approachable for a wider range of users.

3 Methodology

This study is designed as a pilot experiment to explore the feasibility of implementing voice control within a virtual reality environments. Using an Oculus Quest 3 headset coupled with the Wit.ai voice interface, the system enables hands-free navigation and object selection within a simple living room VR environment on Unity.

3.1 Sample

The participants consists of a convenience sample of four participants that include my partner, their mother, and my two children. Although the sample does not represent a broad unbiased population, it provides initial insights into the usability of the voice-controlled VR interface within different age groups.

3.2 Apparatus

The primary hardware component of this study is the Oculus Quest 3. This system is an all-in-one VR headset developed by Meta. It features a high-resolution display with a per-eye resolution of approximately 2064 x 2208 pixels and a wide field of view. The headset comes equipped with inside-out tracking via integrated cameras, allowing for precise head tracking. Additionally, the Oculus Quest 3 is equipped with a built-in microphone for voice command input.

The VR system was developed in Unity with voice input processed using Wit.ai. Commands such as "forward five steps" were parsed into actionable tokens. Movement in the virtual space was executed in increments to match the number of steps dictated in the command. Directional control was achieved using head orientation: if a participant turned their head beyond a threshold angle of 45 degrees, either left or right, their forward-facing direction was reoriented accordingly. This removed the need for a joystick-based turning and made the experience fully hands-free.

3.3 Independent Variable

The independent variable is the implementation of voice commands through the Wit.ai interface. These commands were integrated into the Unity VR system and included actions such as "forward", "forward 5 steps", "back" and "back 5 steps." These were used exclusively to control movement within the virtual environment.

3.4 Dependent Variables

The dependent variables were the performance and user experience outcomes of the system:

- `Task Completion Time`: The quantitative measurement in seconds, this records how long it takes participants to reach the end of the maze.
- `Error Rate`: A quantitative measurement that tracks the number of missed voice commands
- `User Feedback`: A qualitative measurement collected upon completion of the tasks, includes ratings on ease-of-use, satisfaction, and overall experience with the voice controlled VR system.

3.5 Design And Procedure

- `Introduction And Briefing`: The participants were given a brief overview of the study's purpose. They were instructed on how to wear the Oculus Quest 3 and provided with a short demonstration explaining how head movements coupled with voice commands allow for navigation and interaction.
- `Experimental Task`: Participants started at a predefined location within the virtual maze. They were instructed to navigate through the maze until they reach the exit, using head movement, to orient direction, while speaking the command "forward five steps". To ensure reliability, each participant completed multiple trials of the navigation and interaction tasks.
- `Data Analysis`: The dependent variables were collected manually by the researcher, error amount during each task and time to complete each task. After completion of the trials, user feedback was asked via survey consisting of ease-of-use, satisfaction, and overall experience with the voice controlled VR system.

4 Results and Findings

This section presents the quantitative metric data and the qualitative feedback gathered during the pilot study.

4.1 Task Completion Time, Error, and Commands

Participants completed the maze navigation tasks using head movements and voice commands with varying levels of efficiency and accuracy. Table 1 shows the average completion times and number of missed or unrecognized commands for each participant.

Table 1. Average task time and error count by participant

Participant	Avg. Time (s)	Avg. Errors	Avg. Total commands	Recognition Rate
P1	5:18	16	59	72.9%
P2	7:59	23	84	72.6%
P3	5:08	9	58	84.5%
P4	3:29	4	34	88.2%

The recognition rate was calculated using the following formula:

$$\text{Recognition Rate (\%)} = \left(1 - \frac{\text{Errors}}{\text{Total Commands}}\right) \times 100$$

Despite expectations that younger users would perform better, results showed the oldest participant (p4) completed the task the fastest with the fewest errors and fewer commands needed overall. In contrast, p2 issued the most commands (84) and also had the highest error rate (23), leading to the slowest completion time. This inverse relationship between age and error frequency suggests that clarity and consistency in voice input, rather than age, were more influential on system performance.

4.2 Command Recognition and System Performance

Recognition rates demonstrated the relationship between speech clarity and system performance. Older participants (p3 and p4) had higher recognition rates (84.5% and 88.2%), while younger participants (p1 and p2) had lower rates (72.9% and 72.6%). These results suggest that recognition accuracy is dependent on consistent speech patterns and structured command phrasing.

While the recognition rates varied by participant, another key factor influencing the performance was the responsiveness and speed of the natural language processing system Wit.ai. The system needed to not only understand the spoken commands but respond in a timely manner. In many of the trials, this delay would frustrate the situation when the participant is saying the commands multiple times with the wrong intent happening. This finding reinforces the idea that voice interfaces are only as effective as the system's ability to interpret commands quickly and accurately. While AI tools like Wit.ai can handle natural language well, designing for VR requires optimizing for both understanding and speed to avoid disrupting the user.

4.3 Participant Feedback

Participants where asked to complete a survey with 7 scalable questions and 3 optional open feedback questions.

Table 2. All participants used a scale (1 = strongly disagree, 10 = strongly agree) to rate their experience.

Question	Avg. Score
1. How easy was it to learn the voice commands?	7.25
2. How well did the system understand your voice commands?	6.25
3. How responsive was the system after you gave a voice command?	7
4. how natural did the movement feel when using step-based voice commands?	7.25
5. How accessible did you find the VR experience for someone with mobility impairments?	8.75
6. How comfortable was it to use voice instead of physical input?	7.75
7. How satisfied where you with overall experience?	7.75

The participant feedback strongly supports the system's usability and potential for accessible VR interaction. Scores were especially high for accessibly and comfort, although one participant did mention dizziness while performing the test. Additionally, users complained of delayed responsiveness, and accuracy indicating refinements areas for this and future systems.

5 Discussion

The results of this study highlight several important findings regarding the usability and accessibility of a voice-controlled VR interface, particularly for users across different age groups. Contrary to expectations, age was not a limiting factor in successful system usage. In face, the oldest participant (age62) outperformed all others in task completion time, error rate, and recognition accuracy. this suggests that speech clarity and consistent phrasing had a more significant impact on performance that age or familiarity with VR.

Another insight was the impact of the response time and AI interpretation delay. In some trials, recognition lag led to user frustration, particularly when a command was misheard or delayed. This highlights the need for natural language to accurately interpret and invoke the command with little to no delay after the user speaks. To be a viable real-time application, voice controlled VR system must balance natural input with technical responsiveness. The most inclusive systems fail if it causes repeated misfires or delays that break immersion. In future iterations, systems could benefit from:

- Local/offline processing to reduce latency.
- Visual or audio feedback confirming command recognition.
- Adaptive learning models that better handle child and non-native speaker input.

6 Conclusion

This study demonstrates the potential of voice controlled navigation systems to improve accessibility in virtual environments. Developed on Unity for the Oculus Quest 3, using head movement to turn, and integrating Wit.ai for natural language interpretation. This hands-free step-based navigation system offers an alternative to traditional controller-based systems, especially for those who have limited motor functions.

Despite a small and diverse sample, the results showed that speech clarity and consistent phrasing had a great impact on performance. Older participants outperformed younger ones in both recognition accuracy and task completion time. Recognition rates ranged from 72.6% to 88.2%, demonstrating the feasibility of natural language command in VR.

However, the system's effectiveness was limited by response delays and recognition inconsistencies, especially when phrasing varied or input was unclear. These issues require faster and a more accurate AI processing. Future work should involve a broader and more diverse pool of participants. Incorporating adaptive feedback and local speech processing could further enhance usability and accessibility. By reducing reliance on physical controllers and enabling natural interaction, this research contributes to more inclusive virtual environments that prioritize user diversity.

References

[1] Farhan Aslam and Richard Zhao. 2024. *Voice-Augmented Virtual Reality Interface for Serious Games.* https://cspages.ucalgary.ca/~richard.zhao1/publications/2024cog-voice_augmented_VR_interface.pdf#:~:text=Diverse%20approaches%20to%20VR%20interactions,Among Last accessed: March 12, 2025.

[2] Kilian Brachtendorf, Benjamin Weyers, and Daniel Zielasko. 2020. Towards Accessibility in VR - Development of an Affordable Motion Platform for Wheelchairs. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW).* 291–292. doi:10.1109/VRW50115.2020.00064

[3] Anderton C., Creed C., Sarcar S., and Theil A. 2024. *From Teleportation to Climbing: A Review of Locomotion Techniques in the Most Used Commercial Virtual Reality Applications.* https://www.tandfonline.com/doi/full/10.1080/10447318.2024.2372151#abstract

[4] Pedroza-Santiago E.A, Quiroz-Ibarra J.E., Bojorges-Valdez E., and Padilla-Castañeda M.A. 2025. *Comparison of Manual, Automatic, and Voice Control in Wheelchair Navigation Simulation in Virtual Environments: Performance Evaluation of User and Motion Sickness.* https://www.mdpi.com/1424-8220/25/2/530

[5] Jan Hombeck, Henrik Voigt, and Kai Lawonn. 2024. Voice user interfaces for effortless navigation in medical virtual reality environments. *Computers and Graphics* 124, 104069 (November 2024). doi:10.1016/j.cag.2024.104069

[6] University of South Carolina. n.d. *Virtual Environments Accessibility Guidelines.* https://sc.edu/about/offices_and_divisions/cte/teaching_resources/virtual_environments/ve_accessibility_guidelines/#:~:text=Virtual%20Environments%20Accessibility%20Guidelines%20,voice%20commands%2C%20keyboards%2C%20gestures%2C

[7] Jacob O. Wobbrock, Rachel L. Franz, and Melanie Kneitmix. 2024. *Improving the accessibility of virtual reality for people with motor and visual impairments.* https://faculty.washington.edu/wobbrock/pubs/chi-24.02.pdf

Received January 2025 ; revised March 2025

README.txt

Video for YouTube -
Prototype- https://www.youtube.com/watch?v=0h8xNo-Mfvs
Code- https://www.youtube.com/watch?v=mlp9b9p7GUs

overleaf LaTex -https://www.overleaf.com/read/pqwmbmwcyqms#15a54f

references:

meta-Voice SDK - https://assetstore.unity.com/packages/tools/integration/meta-voice-sdk-immersive-voice-commands-264555

manual for VR Template - https://docs.unity3d.com/Packages/com.unity.template.vr@6.1/manual/index.html