Final evaluation report and reflection

Objective and Validation Metrics

The main goal of this evaluation was to test how easily participants could understand and perform the three hand tracking interaction tasks in Discord XR: Galaxy Social Hub. This prototype focused on using natural hand gestures instead of controllers to make social communication in XR more comfortable and realistic. The test aimed to check three main things. First, whether users could recognise what each gesture was meant to do. Second, whether they could complete each task without hesitation or mistakes. Third, whether the experience felt smooth, enjoyable and close to real social interaction. By focusing on these points, the evaluation tried to find out if the prototype could support simple and meaningful social communication through gestures alone. To measure success, I recorded how well users performed each task and how they felt during and after the session. Each task was timed to measure efficiency. The completion rate showed how many participants could finish a task successfully, and the number of errors or times they needed help showed where they faced difficulty. I also wrote short notes about any hesitation or emotional reaction that I observed. At the end, I asked users to rate their satisfaction and share what they liked or wanted to improve. The success criteria were clearly defined before the session. A task was considered successful if at least 80 percent of participants could complete it without help. Efficiency was defined as completing each task within 30 seconds, with confidence shown through positive comments and high satisfaction scores. If users showed confusion, repeated the same gesture several times, or asked for assistance, the task was considered unclear and needed improvement. These measures provided a clear way to judge whether the prototype achieved its goals. Tasks 1 and 2 were expected to reach full completion within short times, showing that users could quickly learn the gestures. Task 3 was designed to test how users would react to a more social action, so success in this task would indicate that the prototype could support natural communication between avatars. By combining performance data and user feedback, this evaluation provided strong evidence of how well the design met its purpose and where it still needed refinement.

Results

| Task ID	Description	Average Time (s)	Total Errors / Assists	Completion (%)	Average Satisfaction Score	Notes
1	Grab planet to enter voice channel	13	0	100		All participants successfully completed this task. They described the explosion effect and smooth transition as interesting and satisfying.
2	Pick up and place down the microphone	7	0	100	4.0	Every participant completed this step easily. They said the microphone felt realistic and liked the ability to throw or move it freely.
3	Poke NPC mute or unmute button	11	1 each	0		None of the participants could finish this task because the function was not yet included. They tried several times and showed confusion.

Table 1: Summary of Task Performance
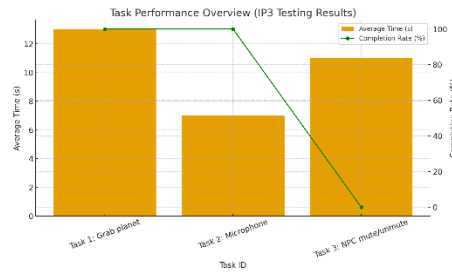
Figure 1: Average Task Completion Time & Rate for Three Tasks

Most participants rated their satisfaction between 4.0 and 4.5 out of 5, saying that the overall interaction was engaging and matched the Discord theme of joining a voice channel. They appreciated the visual effects such as the planet explosion and found the microphone feature fun to explore. Common suggestions included adding multi user participation, NPC reactions and sound feedback when tasks are completed. Several participants also said that they would like to have clearer guidance at the start, such as short text or visual hints showing what can be done in each scene.

Analysis and Insights

The testing results showed both strong and weak points in how users interacted with the prototype. By looking at the task completion, time, and comments, four main findings became clear. These findings together reveal that the prototype successfully demonstrated natural gesture control but still needs stronger feedback and social depth to fully support user confidence.

1. Task Completion and Ease of Use

The first two tasks were finished by all five participants with 100 percent success and no errors. The average time for Task 1 was 13 seconds, and for Task 2 it was only 7 seconds. These short times show that the gestures were clear and easy to perform. Participants did not need much help, which means that hand tracking worked smoothly and the gestures were easy to understand. However, Task 3 was not completed by anyone because the function was missing. The average attempt time was about 11 seconds before users stopped trying. This gap between the first two successful tasks and the final failed one clearly shows that missing features can interrupt the whole test flow and affect user confidence. These results suggest that users can easily learn and adapt to natural hand gestures, but their trust in the system quickly drops when actions lack a clear response or end point.

2. Enjoyment and Realism

All participants said that they enjoyed the first two tasks and that the interactions felt natural and satisfying. The explosion effect when entering a new scene and the realistic microphone movement were often mentioned as highlights. Several users repeated the microphone action just for fun, which shows a high level of engagement. Their average rating is 4 out of 5, also prove that the system was enjoyable to use. These reactions suggest that visual and physical feedback plays an important role in creating a feeling of realism. When gestures match users' expectations, they feel more connected and are willing to explore the environment longer. This pattern reveals that users value playful and responsive design elements that make the virtual world feel alive and rewarding.

3. Problems and Missing Feedback

Although Tasks 1 and 2 were successful, most participants wanted more signs to confirm their actions. For example, after grabbing a planet, they were not sure if the system had fully recognised the gesture until the explosion appeared. A short sound or colour change could make this clearer. The same

problem appeared with the microphone interaction, where users expected vibration or sound feedback when they picked it up. The average completion time shows that users could still perform well, but stronger feedback could make them more confident and faster in future tests. The confusion in Task 3 also shows how important it is to provide feedback even when a function is not yet complete. This finding highlights that clear and immediate feedback is key to maintaining user confidence and reducing hesitation, especially when exploring social gestures in XR.

4. Overall Performance and Future Focus

The overall performance of the prototype was good. The average task time across all tasks was less than 15 seconds, showing that users were efficient when features worked correctly. The completion rate for Tasks 1 and 2 reached 100 percent, which confirms that the hand tracking system is reliable. However, the total success rate across all three tasks dropped to around 67 percent because of the missing third function. The numbers together with user comments show that the concept of using hand gestures for social XR communication is strong, but it still depends on clear guidance and complete functions. The next version should focus on giving users more visible and sound feedback, so that they always know when an action has succeeded or failed. Overall, the results reveal that the foundation of the concept is solid, but the success of social interaction in XR depends on completing every function and creating a seamless feedback loop between user actions and system responses.
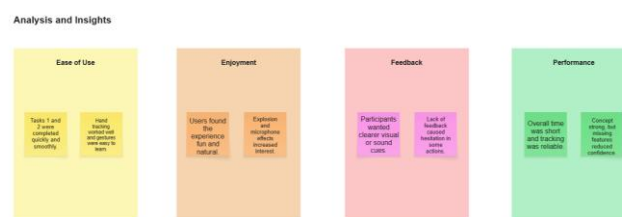

Figure 2: Key Themes Identified in User Evaluation

Evaluation of Aims

The testing goals were mostly achieved and provided a clear understanding of how the prototype performed. The first goal was to check whether users could understand and complete all three gestures easily. This goal was only partly achieved because Tasks 1 and 2 were completed successfully, while Task 3 failed due to the missing function. The second goal, which focused on system comfort and gesture smoothness, was fully achieved. Participants performed all movements naturally and reported no delays or tracking problems. The third goal aimed to see if the experience was engaging and connected well with the Discord theme. This was achieved since users gave positive comments and high satisfaction scores, describing the prototype as fun and realistic. However, the goal of providing clear feedback was only partly achieved because participants still requested stronger visual and sound cues to confirm their actions. Finally, the goal of testing a complete social interaction loop was not achieved because the NPC mute or unmute function was not implemented. Overall, the evaluation confirmed that the prototype demonstrated strong usability and engagement, but it also revealed that completeness and feedback are essential for improving user confidence and social realism in future iterations. The results validated that hand tracking is effective and enjoyable for most users, but they also showed that the absence of full social features left the sense of communication incomplete. Some aspects, such as how users would react to real-time voice interaction, remain uncertain and will require further testing in future versions.

Reflection

Prototype Session Review

The IP3 testing session helped identify both progress and problems. Compared with earlier versions, the hand tracking system ran more smoothly and could correctly detect pinching, grabbing, and placing gestures. Participants said the improved responsiveness and explosion effect made the interaction more enjoyable and helped them understand when they had entered a new scene. These aspects worked well and made the experience feel polished. However, the missing Task 3 stopped users from completing the full interaction loop, which made them uncertain about the goal of the session. The lack of a clear ending also reduced their sense of achievement, even though the first two tasks were smooth. This showed that even when most functions work well, one missing feature can make the overall experience feel incomplete. I learned that testing needs to simulate the whole user journey so participants can understand how each step connects to the next.

Methodological Reflection

The testing plan was clear and workable, combining both quantitative and qualitative methods. These methods worked well for collecting structured results. However, some actions and expressions were missed because it was difficult to take notes while observing participants' reactions. I also noticed that switching between the timer, note sheet, and observation distracted my attention from subtle gestures and emotional responses. If I could repeat the process, I would use voice recording to capture participants' comments and reactions in real time. I would also prepare a short pilot test to make sure all functions, timers, and reset steps work before inviting participants. This way, I can reduce technical errors and collect more accurate evidence about user behaviour and comfort.

Concept Evaluation

This test partly confirmed my design concept. The idea of using natural hand gestures instead of controllers worked well, as all participants could complete grabbing and placing actions easily and described them as intuitive. The microphone feature also helped them feel a sense of control and realism, proving that simple, tactile interactions can make XR environments more engaging. However, the NPC response lacked, users wanted reactions from others or visible changes to confirm that their action had meaning. I learned that good interaction design must balance physical control and social connection, ensuring users always receive meaningful responses to their gestures. This experience also made me realise that adding small emotional cues, such as gestures of acknowledgement or sound reactions, can make interaction feel more natural and alive in XR.

Improvements and Extensions

The findings show several clear directions for improvement. The missing NPC interaction should be added so that users can see a quick and visible response after poking or speaking. This will help them feel that the environment reacts to their presence. The system also needs stronger feedback cues, such as a short sound when the planet is grabbed or a slight glow when the microphone is activated. During testing, users often paused for confirmation, so these small cues could reduce hesitation and improve confidence. The starting scene could also include a simple tutorial panel explaining available gestures, since new users often asked what they could do next. In the future, I would like to explore how different types of visual and sound cues can affect user confidence and timing during interaction, and whether they still help users stay focused when multiple actions happen at the same time. These improvements would not only fix existing issues but also create a more expressive and complete system where users clearly understand what happens after each action.

Appendix

Testing Result IP3

Task 1: Remotely grab a planet representing a voice channel to enter it.
Task 2: Pick up and place down the microphone on the table.
Task 3: Poke another NPC's mute or unmute button.

U1

User ID	Task ID	Success (Y/N)	Errors / Assists	Time (s)	Observation Notes
1	1	Y	0	13	The participant found the remote grabbing interaction and entering a new scene very interesting, and praised the explosion effect that appeared after the grab.
	2	Y	0	7	The participant really enjoyed the microphone interaction and was pleasantly surprised to discover that it could be thrown, so he tried it several times with excitement.
	3	N	1	11	Because the prototype did not include this function, the participant tried several times but failed and felt a bit confused.

Open Questions

How satisfied were you with the overall interaction experience?

If I were to give a score, I would say 4 out of 5. The whole process was engaging and matched well with the Discord theme of joining a voice channel and gathering many people to chat together.

If you could add one new feature, what would it be?

I hope it could become a multi-user prototype where several people can join together, not just interactions with NPCs.

What would you like to improve or change?
First, the interaction in Task 3 could be implemented. Second, since both scenes are very important, they could be further refined and improved.

U2

User ID	Task ID	Success (Y/N)	Errors / Assists	Time (s)	Observation Notes

User ID	Task ID	Success (Y/N)	Errors / Assists	Time (s)	Observation Notes
2	1	Y	0	13	The participant quickly understood how to grab the planet and mentioned that the scene transition felt smooth.
	2	Y	0	7	The participant enjoyed experimenting with the microphone, trying to throw it several times and commenting that it felt very realistic.
	3	N	1	11	

Open Questions

How satisfied were you with the overall interaction experience?

Rated 4.5 out of 5. The whole experience was interesting and matched the theme of joining a Discord voice channel.

If you could add one new feature, what would it be?

Adding background sound or ambient voices would make the world feel more lively.

What would you like to improve or change?
NPCs could have clearer reactions when being poked.

U3

User ID	Task ID	Success (Y/N)	Errors / Assists	Time (s)	Observation Notes
3	1	Y	0	13	Participant showed some hesitation before realising that the planet could be grabbed remotely.
	2	Y	0	7	
	3	N	1	11	Several attempts were made to poke the NPC button, but no clear response occurred, which caused some confusion.

Open Questions

How satisfied were you with the overall interaction experience?

Rated 4 out of 5. The concept was creative and the gestures were easy to understand.

If you could add one new feature, what would it be?

A voice function could be added so that users can actually speak through the microphone.

What would you like to improve or change?
Interactive objects could have clearer visual indicators or tooltips.

U4

User ID	Task ID	Success (Y/N)	Errors / Assists	Time (s)	Observation Notes
4	1	Y	0	13	The explosion effect after entering the new scene was noted as engaging and satisfying.
	2	Y	0	7	The microphone interaction worked smoothly, and the gesture was completed without any difficulty.
	3	N	1	11	Because the prototype did not include this function, the participant tried several times but failed and felt a bit frustrated.

Open Questions

How satisfied were you with the overall interaction experience?

Rated 4 out of 5. The interaction expect task 3 felt natural and the flow was very smooth.

If you could add one new feature, what would it be?

An option to wave or express simple gestures to other avatars could increase realism.

What would you like to improve or change?
Sound feedback for the microphone would make it more engaging

U5

User ID	Task ID	Success (Y/N)	Errors / Assists	Time (s)	Observation Notes
5	1	Y	0	13	The participant completed the first two tasks smoothly and commented that the interactions felt simple and responsive.
	2	Y	0	7	
	3	N	1	11	

Open Questions

How satisfied were you with the overall interaction experience?

Rated 3.5 out of 5. The interaction was enjoyable, but some gestures did not respond as expected.

If you could add one new feature, what would it be?

Allowing multiple users to join and interact together would make the experience more social.

What would you like to improve or change?
Clearer instructions could be given at the start, and every gesture should include visual or sound feedback.

References

[AI-1 | All Scripts]

Claude (Anthropic AI), used to developed the code's structure. **Retrieved via claude.ai, Last Accessed: 24/10/2025.**

[AI-2 | Report grammar editing]
ChatGPT (OpenAI), *used to correct grammar and improve sentence structure in the written report.* **Retrieved via chat.openai.com, Last Accessed: 29/10/2025.**

Statement of Original Work

I certify that this project is my own original work, except where otherwise indicated. ChatGPT was used to assist with grammar correction and sentence structure improvements in the written report. Claude was used to develop the structure with some sections of the Unity C# code. The design concept, testing plan, evaluation, analysis, and the majority of the implementation are entirely my original work. Importantly, all AI-assisted code was reviewed, explained in my own words, and adapted to fit the context of this project, ensuring that it reflects my own understanding and integration.