Evaluation - Interactive Prototype 1

Objective and Validation Metrics

The main objective of this evaluation was to test three core interactions within the Discord XR Galaxy Social Hub, which is a redesign of Discord into an XR environment where planets represent voice channels and users interact through avatars in a galaxy-like virtual space. The three interactions were: joining a channel by rotating and dragging planets, sitting and changing seats with the F key, and muting or unmuting other users by clicking the zipper icon. The evaluation aimed to determine whether these interactions were discoverable, efficient, and easy to use. To determine success, four criteria were applied. Task completion time in seconds was measured to evaluate efficiency. The number of errors observed provided a measure of accuracy and clarity. The System Usability Scale (SUS) was used to capture participants' overall impression of usability, generating scores from 0 to 100. Finally, qualitative feedback from Think Aloud protocols and a post-test open-ended question was collected to provide insight into user perceptions, frustrations, and positive reactions. Together, these measures allowed for both quantitative assessment of performance and qualitative interpretation of user experience.

Results

The user testing involved five participants who completed three main tasks. The quantitative results showed that all participants finished their tasks within the expected time caps. For Task 1, joining a channel by rotating and dragging planets, the average completion time was 21.6 seconds with approximately one error per user. For Task 2, sitting and changing seats with the F key, the average time was 25.6 seconds and all participants successfully completed the task with two errors. Task 3, muting and unmuting other users by clicking the zipper icon, was the fastest, with an average completion time of 14.2 seconds and two errors. The SUS scores ranged from 67.5 (grade D) to 85.0 (grade A), with a mean score of 75.5. This indicates that the prototype reached an acceptable level of usability but also suggests room for improvement.

The qualitative results highlighted several recurring issues. Many participants described the rotation and dragging interaction as awkward and confusing, with some noting that they could not tell whether the rotation was working correctly. The zipper mute icon was consistently identified as too small and visually unclear, with several users suggesting a larger or more recognisable symbol. In contrast, participants responded positively to playful features such as using the F key to sit and change seats, which they considered fun and engaging. Suggestions from users included adding clearer exit options and providing stronger visual cues when joining channels or muting users. These observations provided valuable insights into both strengths and weaknesses of the prototype.

Task	Avg. Time(s)	Total Error	Completion Rate (%)	Notes
Task 1: Join Channel	21.6	5	100	Users found rotation/dragging awkward
Task 2: Sit & Change Seats	25.6	2	100	Users enjoyed the F key feature
Task 3: Mute/Unmute	14.2	2	100	Icon too small/unclear

Table 1: Summary of Task Performance

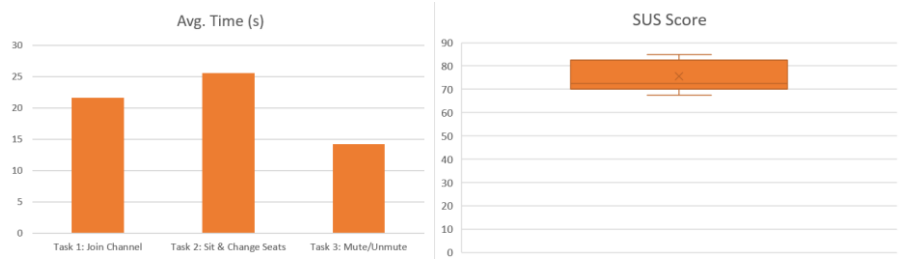


Figure 1 & 2: Average Task Completion Time for Three Tasks & Distribution of SUS Scores Across Participants

Analysis and Insights

The evaluation results can be interpreted through four main themes. Discoverability was very crucial since participants struggled with the planet rotation and dragging interaction. Although all users eventually joined a channel, their repeated comments about awkwardness and uncertainty demonstrated that the interface lacked natural cues to guide user actions. Clarity was another recurring theme, particularly for the mute function. Even though Task 3 was completed quickly, users consistently reported that the zipper mute icon was too small and unclear. This shows that efficiency alone does not ensure usability if the visual design of controls is weak, as poor affordances reduce confidence and increase frustration. Engagement emerged as a positive theme, with several participants expressing enjoyment when discovering the F key function to sit and change seats. This playful element encouraged deeper interaction and suggests that social XR environments benefit from a balance of functionality and enjoyable features. Finally, Feedback was a strong demand across all tasks. Participants wanted clear signals when they had joined a channel, muted another user, or changed seats. The lack of obvious visual or auditory confirmation forced users to rely on trial and error, slowing task performance and reducing satisfaction. These themes reveal that while the prototype supported successful task completion, its usability and user experience can be significantly improved through clearer affordances, better feedback, and attention to visual clarity, while still maintaining engaging and playful features.

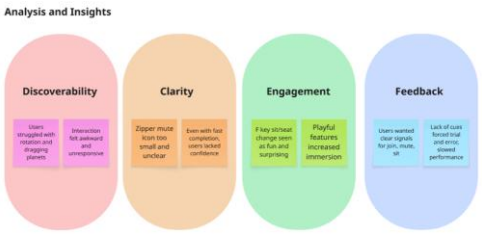


Figure 3: Key Themes Identified in User Evaluation

Evaluation of Aims

The testing objectives were partly achieved and provided useful evidence about the prototype's current strengths and weaknesses. The assumption that users could complete the three interactions was validated, as all participants successfully managed to join channels, sit and change seats, and mute or unmute another user within the given time limits. This confirmed that the system design was functional and supported the intended workflow. However, the assumption that these interactions would be efficient and easy to understand was only partially validated. While some participants completed the tasks smoothly, others hesitated or repeated actions, especially during rotation and dragging, which showed inconsistency in usability. The assumption that the zipper mute control would be intuitive was clearly invalidated, as most users described the icon as confusing, too small,

and not visually distinct enough from the rest of the interface. These findings indicate that the prototype achieved fundamental usability and confirmed the feasibility of the concept, but significant refinements are required to achieve higher-level goals such as smooth discoverability, stronger affordances, and consistent clarity across all interactions.

Concept Iteration

Through the process, the value of combining quantitative and qualitative methods has been evaluated. First, the planet rotation and dragging mechanism should be improved, potentially by using a central axis or an alternative interaction technique, so that the movement feels smoother, more predictable, and easier for users to learn. Second, the mute function should be updated with a larger and more recognisable icon, ideally adopting a conventional speaker symbol or a similar clear indicator, to ensure that users can immediately identify its purpose without confusion. Third, exit options should be made more visible and easily accessible, as participants expressed uncertainty about how to leave channels or cancel actions, and clearer navigation controls would reduce frustration. Finally, stronger visual and animated feedback should be introduced to confirm when an action has been successfully completed, such as colour changes, highlight effects, or short audio cues for joining, muting, or sitting. These improvements will directly address the weaknesses identified during testing and make the system more intuitive, reliable, and engaging. In addition, preserving playful elements, such as the seating change feature, will help maintain user interest and enjoyment while the overall usability is strengthened.

Reflection

Reflecting on the process, this evaluation highlighted the value of combining quantitative and qualitative methods. While task times and error rates provided objective evidence that users could complete tasks, they did not reveal the underlying frustrations or uncertainty. The Think Aloud protocols and post-test comments filled this gap, showing where interactions lacked discoverability and clarity. This confirms that both types of data are essential to understanding usability in XR systems.

The testing plan was generally successful, but there were areas for improvement. A few participants required additional explanation at the beginning of the test, which suggests that task instructions should be clearer and more concise in future studies. Introducing visual prompts or step-by-step task cards may improve participant understanding and reduce hesitation. The organisation of qualitative notes could also be improved by using a structured logging template to capture observations consistently and reduce the risk of missing important details.

At a broader level, this evaluation reinforced the importance of iteration in the design process. The prototype demonstrated its intended functions, but the findings showed that discoverability, feedback, and affordances require further refinement. Designing for XR is not only about technical implementation but also about ensuring that interactions feel natural and intuitive. For the next iteration, I will focus on redesigning the planet rotation and mute icon, while maintaining engaging elements such as seating changes that users enjoyed.

Finally, this evaluation reinforced the principle that assumptions must always be validated with real users. Even though I expected the zipper mute control to be straightforward, participants found it confusing. This experience reminded me that user testing exposes gaps that are difficult to predict during design. For future work, I will apply this knowledge to improve this prototype and strengthen my overall approach to XR design, evaluation, and iteration.

Appendix

Quantitative Results

User	T1_time_s	T1_errors	T2_time_s	T2_errors	T3_time_s	T3_errors
U1	24	1	23	0	27	2
U2	13	0	19	1	12	0
U3	28	2	32	0	9	0
U4	22	1	28	0	16	0
U5	21	1	26	1	7	0

User	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	SUS score	Letter Grade
U1	5	2	3	4	5	1	4	2	3	2	72.5	C
U2	4	2	3	3	5	2	4	3	2	1	67.5	D
U3	3	1	4	4	5	2	4	2	3	1	72.5	C
U4	5	1	4	2	4	1	5	2	4	2	85.0	A
U5	5	2	4	1	3	2	4	1	4	2	80.0	B

Qualitative Notes

User	Think Aloud	End question-What is one thing you would fix?
U1	• I keep forgetting I have to hold to join. • It'd be nicer if everyone started unmuted and showed up green.	For task 2, after sitting, face other users' avatars.
U2	• Task 2 is a bit unclear, I don't know if I can I sit here or not. • A little hint would help. • Also, the zipper mute is tiny and hard to click.	Use a central axis for rotating planets; left-click to rotate the axis instead of right-dragging planets.
U3	• I can't tell if the rotation is working, it feels unresponsive. • Please add an Exit for task 2 and 3. • That small black bar isn't clear. • A speaker icon would help.	Provide clearer cues for how joining occurs in Task 1.

U4	• Rotating to pick a planet feels awkward. • I just found out F lets me sit, stand, and switch seats, which is cool.	Allow avatar to fly to the planet; dragging the planet itself can be confusing.
U5	• Right-click dragging is a bit fiddly. • Movement feels good though. • The zipper should be bigger.	Make zipper control larger a

The University of Queensland 8

Yesterday 4:26

The University of Queensland 7

Yesterday 5:35

The University of Queensland 6

Yesterday 8:03

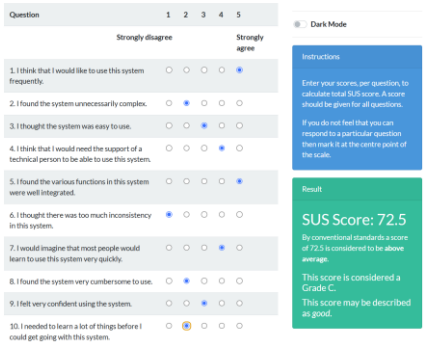
The University of Queensland 5

Yesterday 3:06

The University of Queensland 4

Yesterday 5:30

User 1 SUS



Question	1	2	3	4	5
	Strongly disagree				Strongly agree
1. I think that I would like to use this system frequently.	○	○	○	○	◉
2. I found the system unnecessarily complex.	○	◉	○	○	○
3. I thought the system was easy to use.	○	○	◉	○	○
4. I think that I would need the support of a technical person to be able to use this system.	○	○	○	◉	○
5. I found the various functions in this system were well integrated.	○	○	○	○	◉
6. I thought there was too much inconsistency in this system.	◉	○	○	○	○
7. I would imagine that most people would learn to use this system very quickly.	○	○	○	◉	○
8. I found the system very cumbersome to use.	○	◉	○	○	○
9. I felt very confident using the system.	○	○	◉	○	○
10. I needed to learn a lot of things before I could get going with this system.	○	◉	○	○	○

Dark Mode

Instructions

Enter your scores, per question, to calculate total SUS score. A score should be given for all questions.

If you do not feel that you can respond to a particular question then mark it at the centre point of the scale.

Result

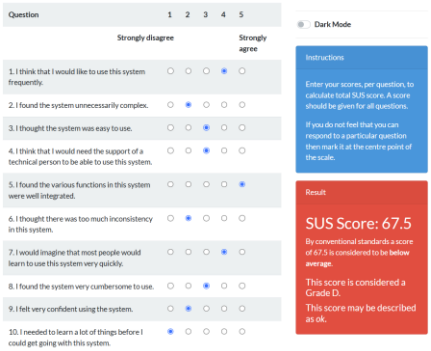
SUS Score: 72.5

By conventional standards a score of 72.5 is considered to be **above average**.

This score is considered a Grade C.

This score may be described as *good*.

User 2 SUS



Question	1	2	3	4	5
	Strongly disagree				Strongly agree
1. I think that I would like to use this system frequently.	○	○	○	◉	○
2. I found the system unnecessarily complex.	○	◉	○	○	○
3. I thought the system was easy to use.	○	○	◉	○	○
4. I think that I would need the support of a technical person to be able to use this system.	○	○	◉	○	○
5. I found the various functions in this system were well integrated.	○	○	○	○	◉
6. I thought there was too much inconsistency in this system.	○	◉	○	○	○
7. I would imagine that most people would learn to use this system very quickly.	○	○	○	◉	○
8. I found the system very cumbersome to use.	○	○	◉	○	○
9. I felt very confident using the system.	○	◉	○	○	○
10. I needed to learn a lot of things before I could get going with this system.	◉	○	○	○	○

Dark Mode

Instructions

Enter your scores, per question, to calculate total SUS score. A score should be given for all questions.

If you do not feel that you can respond to a particular question then mark it at the centre point of the scale.

Result

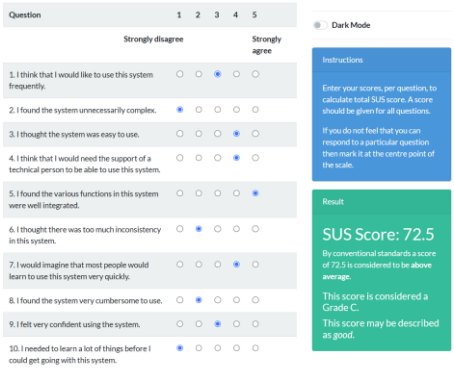
SUS Score: 67.5

By conventional standards a score of 67.5 is considered to be **below average**.

This score is considered a Grade D.

This score may be described as *ok*.

User 3 SUS



Question	1	2	3	4	5
	Strongly disagree				Strongly agree
1. I think that I would like to use this system frequently.	○	○	◉	○	○
2. I found the system unnecessarily complex.	◉	○	○	○	○
3. I thought the system was easy to use.	○	○	○	◉	○
4. I think that I would need the support of a technical person to be able to use this system.	○	○	○	◉	○
5. I found the various functions in this system were well integrated.	○	○	○	○	◉
6. I thought there was too much inconsistency in this system.	○	◉	○	○	○
7. I would imagine that most people would learn to use this system very quickly.	○	○	○	◉	○
8. I found the system very cumbersome to use.	○	◉	○	○	○
9. I felt very confident using the system.	○	○	◉	○	○
10. I needed to learn a lot of things before I could get going with this system.	◉	○	○	○	○

Dark Mode

Instructions

Enter your scores, per question, to calculate total SUS score. A score should be given for all questions.

If you do not feel that you can respond to a particular question then mark it at the centre point of the scale.

Result

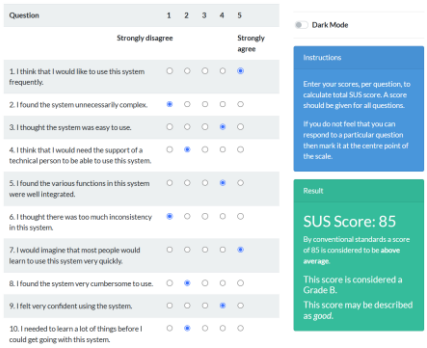
SUS Score: 72.5

By conventional standards a score of 72.5 is considered to be **above average**.

This score is considered a Grade C.

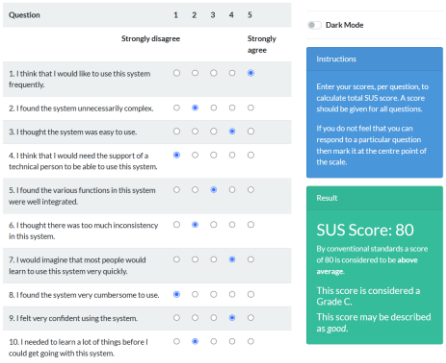
This score may be described as *good*.

User 4 SUS



Question	1	2	3	4	5
	Strongly disagree				Strongly agree
1. I think that I would like to use this system frequently.	○	○	○	○	◉
2. I found the system unnecessarily complex.	◉	○	○	○	○
3. I thought the system was easy to use.	○	○	○	◉	○
4. I think that I would need the support of a technical person to be able to use this system.	○	◉	○	○	○
5. I found the various functions in this system were well integrated.	○	○	○	◉	○
6. I thought there was too much inconsistency in this system.	◉	○	○	○	○
7. I would imagine that most people would learn to use this system very quickly.	○	○	○	○	◉
8. I found the system very cumbersome to use.	○	◉	○	○	○
9. I felt very confident using the system.	○	○	○	◉	○
10. I needed to learn a lot of things before I could get going with this system.	○	◉	○	○	○

User 5 SUS



Question	1	2	3	4	5
	Strongly disagree				Strongly agree
1. I think that I would like to use this system frequently.	○	○	○	○	◉
2. I found the system unnecessarily complex.	○	◉	○	○	○
3. I thought the system was easy to use.	○	○	○	◉	○
4. I think that I would need the support of a technical person to be able to use this system.	◉	○	○	○	○
5. I found the various functions in this system were well integrated.	○	○	◉	○	○
6. I thought there was too much inconsistency in this system.	○	◉	○	○	○
7. I would imagine that most people would learn to use this system very quickly.	○	○	○	◉	○
8. I found the system very cumbersome to use.	◉	○	○	○	○
9. I felt very confident using the system.	○	○	○	◉	○
10. I needed to learn a lot of things before I could get going with this system.	○	◉	○	○	○

References

[AI-1 | PlanetInteraction.cs] ***"Unity drag and transition logic"*, Claude (Anthropic AI). Retrieved via claude.ai, Last Accessed: 28/08/2025.**
[AI-2 | SimplePlayer.cs] ***"Mute toggle and visual feedback logic"*, Claude (Anthropic AI). Retrieved via claude.ai, Last Accessed: 28/08/2025.**
[AI-3 | ZipperDragSystem.cs] ***"Unity drag-to-mute interaction design"*, Claude (Anthropic AI). Retrieved via claude.ai, Last Accessed: 28/08/2025.**
[AI-4 | Report grammar editing] **ChatGPT (OpenAI), *used to correct grammar and improve sentence structure in the written report*. Retrieved via chat.openai.com, Last Accessed: 03/09/2025.**

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 provide assistance with some sections of the Unity C# code. All AI-assisted contributions are clearly identified with in-code comments and cross-referenced in the References section. 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.