國立臺灣大學電機資訊學院資訊工程學系

碩士論文

Department of Computer Science and Information Engineering

College of Electrical Engineering and Computer Science

National Taiwan University

Master Thesis

桌面版元宇宙平台中的使用者互動設計

User Interaction Design for Desktop Metaverse

許博翔

Po-Hsiang Hsu

指導教授: 洪一平 博士

Advisor: Yi-Ping Hung Ph.D.

中華民國 112 年 8 月

August, 2023

doi:10.6342/NTU202301630

國立臺灣大學碩士學位論文
口試委員會審定書
MASTER'S THESIS ACCEPTANCE CERTIFICATE
NATIONAL TAIWAN UNIVERSITY

桌面版元宇宙平台中的使用者互動設計

User Interaction Design for Desktop Metaverse

本論文係許博翔君（學號 R10922017）在國立臺灣大學資訊工程學系完成之碩士學位論文，於民國 112 年 7 月 26 日承下列考試委員審查通過及口試及格，特此證明。

The undersigned, appointed by the Department of Computer Science and Information Engineering on 26 July 2023 have examined a Master's thesis entitled above presented by HSU, PO-HSIANG (student ID: R10922017) candidate and hereby certify that it is worthy of acceptance.

口試委員 Oral examination committee:

（指導教授 Advisor）	賴尚宏	
	鄭文皇	王元凱

系主任/所長 Director: 洪士灝



致謝

首先，我要感謝一路以來不斷關心我的家人，尤其是三不五時打電話關心我心情、論文進度的爸媽，雖然有時會覺得嘮叨，但同時也因為有人關心、在意，而感受到幸福。他們也總在我意志消沈的時刻，帶給我溫暖、光亮。

我也要感謝我的指導教授洪一平老師。在這兩年，接觸到幾個案子，我累積了不少實作經驗，也學會跟其他人溝通。在一次次討論論文的過程中，老師總是耐心地指導、表達建議，讓我能把這個研究做得更好。老師時而天馬行空的想法，往往能帶給我更遼闊的視野，看見那些原本看不見的事物。

接著，也要感謝在實驗室認識的好朋友們。感謝一起為畢業而努力的同學們筱晴、宜儒、泂杌，讓我在這一段路並不是孤軍奮戰。感謝鴻明，雖然沒有一起做研究，但也在過程中給了我不少建議和想法。感謝宥杏，就算人在德國，也願意傾聽我心中的負能量。感謝冠廷，時常陪我討論、解決許多難題。感謝敬媛、張琪、嘉文、允玟、舒盟、靖婷、世彬、呂靜、英杰、品文，一次次的晚餐、唱歌、旅行、慶生，讓我的實驗室生活變得更充實更有趣。感謝義峰，研究過程中提供我不少協助，讓我能更快把事情做好。

最後，我要感謝那個就算遇到很多困難還有很大壓力卻還是不斷努力的自己。如果在某個時刻停下腳步，就不會考進台大、繼續在台大念研究所，然後寫出這篇論文。

doi:10.6342/NTU202301630



摘要

近年來，隨著網路以及 AR, VR 和 MR 技術的進步，元宇宙正蓬勃發展。在本文中，我們聚焦於桌面版元宇宙中的互動設計，並試圖強化使用者在其中的體驗。我們比較了幾個知名的元宇宙平台，並實作了一個包含聊天功能、虛擬替身以及支援不同視角（第一及第三人稱）的元宇宙平台。在過程中我們發現了幾個值得深入探討的主題，並以此進行使用者研究。實驗結果顯示了「個別關閉對不同使用者說話的麥克風和來自不同使用者聲音」以及「根據與本地使用者之間距離調整遠端使用者說話音量」這兩項功能的重要性，也得到了使用者在第三人稱視角下對於固定視角有強烈偏好的結果。基於這些結果，我們提出了助於提升桌面版元宇宙使用者體驗的幾個設計建議。

關鍵字：元宇宙、互動設計、多人互動、語音聊天、相機視角

doi:10.6342/NTU202301630



Abstract

Nowadays, with the advancement of network and AR, VR, and MR technologies, the popularity of the metaverse has flourished a lot. In this paper, we explored the desktop metaverse's interaction designs and proposed enhancements to elevate user experience. We examined and compared popular metaverse platforms, and implemented our metaverse application with features including chat, avatar, and different perspectives (first-person and third-person). We uncovered some points for further investigation during the process and conducted a user study on these points. The results highlight the importance of individual controls for muting the microphones to others and the sounds of others, as well as the significance of adjusting remotes' sound volume based on the distance to the local user. Additionally, our findings reveal a notable preference for the fixed view over the free view in the third-person mode. Based on these insights, we present critical recommendations for optimizing the interaction design for the desktop metaverse.

Keywords: Desktop metaverse, interaction design, multi-user interaction, voice chat, camera perspective

doi:10.6342/NTU202301630



Contents

Page

口試委員審定書 i

致謝 ii

摘要 iii

Abstract iv

Contents v

List of Figures vii

List of Tables ix

Chapter 1 Introduction 1

Chapter 2 Related Works 4

Chapter 3 System Design 7

3.1 System Architecture 7

3.2 User Interface 8

3.3 Components & Features 10

3.3.1 Text Chat 10

3.3.2 Voice Chat 12

3.3.3 Avatar 13

3.3.4 Perspective 15

doi:10.6342/NTU202301630

Chapter 4 User Study **17**

4.1 User Study 1 17

4.1.1 Objectives 17

4.1.2 Setups 18

4.1.3 Process 19

4.2 User Study 2 20

4.2.1 Objectives 20

4.2.2 Setups 20

4.2.3 Process 20

Chapter 5 Results **22**

5.1 User Study 1 22

5.1.1 Voice Chat 22

5.1.2 Perspectives 25

5.2 User Study 2 30

Chapter 6 Discussion **33**

6.1 Voice Chat 33

6.2 Perspective 34

6.3 System Limitation 34

Chapter 7 Conclusion **36**

Chapter 8 Future Work **38**

References **40**

doi:10.6342/NTU202301630



List of Figures

3.1 Connecting architecture of our system. 8
3.2 User interface of our system 9
3.3 User interface for text chat 11
3.4 Example of individual sound control 14
3.5 Free view vs. fixed view 16

4.1 Our system with user study setups. 18

5.1 Results of ”I would like to mute the sounds of others separately.” in User Study 1 (1: strongly disagree; 5: strongly agree) 23
5.2 Results of ”I would like to mute my microphones to others separately.” in User Study 1 (1: strongly disagree; 5: strongly agree) 24
5.3 Results of ”I can feel myself wandering in the space when moving my avatar.” (1: strongly disagree; 5: strongly agree) 26
5.4 Results of ”I feel I’m talking with other users face-to-face.” (1: strongly disagree; 5: strongly agree) 26
5.5 Results of ”I can clearly know my position in the virtual space.” (1: strongly disagree; 5: strongly agree) 27
5.6 Results of ”I like to explore the space with this perspective.” (1: strongly disagree; 5: strongly agree) 28
5.7 Results of ”Which perspective do you like the most?” 28
5.8 Results of ”Do you like free or fixed view more?” 29
5.9 Results of ”I would like to mute the sounds of others individually.” in User Study 2 (1: strongly disagree; 5: strongly agree) 31

doi:10.6342/NTU202301630

5.10 Results of ”I would like to mute my microphones to others individually.” in User Study 2 (1: strongly disagree; 5: strongly agree) 31
5.11 Results of ”In which room did you have a better voice chat experience?” . 32

doi:10.6342/NTU202301630



List of Tables

2.1 Comparison of some popular metaverse platforms and our implementation 6

doi:10.6342/NTU202301630



Chapter 1 Introduction

In recent years, with the advancements of VR, AR, and MR technologies, the metaverse appeared and has flourished a lot. *"The metaverse is the next evolution in social connection and the successor to the mobile internet,"* stated *Meta*, one of the leading companies dedicated to the metaverse. The metaverse allows people not physically in the same place to interact, chat, play, learn, and even attend virtual concerts together, creating an experience that resembles being physically present in the same location [8].

Metaverse platforms, often presented in virtual reality (VR), can be operated either on a head-mounted display (HMD) or a personal computer (PC). In this paper, we refer to them as "HMD Metaverse" and "Desktop Metaverse," respectively, as VR is commonly classified into "HMD VR" and "Desktop VR" depending on the type of device used [5–7, 11, 15]. With HMD, users are surrounded by a virtual world, controlling the system with controllers or hands; the content of the desktop metaverse is displayed on a 2D screen, and users control the system with a mouse or a keyboard. Although the former may offer more immersion, the latter is more accessible due to the popularity of computers. According to Verified Market Research [13], the desktop metaverse represents the highest revenue share, implying enormous popularity. Thus, in this study, we are focusing on the design of the interactive features for the desktop metaverse.

doi:10.6342/NTU202301630

We started by comparing some of the popular metaverse platforms, including *Decentraland* [3], *Sandbox* [12], *Oncyber* [9], *Spatial* [10], and *VRChat* [14]. These platforms run on PC, while *VRChat* also runs on HMD. We focused on features like voice chat, the interaction between avatars, and the user's perspective. Then, we found that the interaction designs vary from one to another. For example, in terms of voice chat, *Decentraland* and *VRChat* allow users to mute specific users, while *Oncyber* and *Spatial* only offer the option to mute everyone. Regarding perspectives, *Decentraland* and *Sandbox* only have a third-person view, while the others allow unrestricted switching between the two modes. Despite the diversity of interaction designs, no previous work investigated design suggestions. Therefore, we want to determine which design choices are better for the users.

In this paper, our metaverse platform is developed and implemented based on *akaVerse* [1], a web-based NFT showroom. Presently, users can only explore *akaVerse* individually. Our objective was to create a metaverse platform that enables collaborative visits to exhibitions, where hosts or teachers can provide introductions or lectures. At the same time, guests or students can engage in private discussions. To achieve this, we incorporated several interactive elements into the virtual space, including text chat, voice chat, avatars, and the ability to switch between first and third-person perspectives. During the implementation, we identified specific user experience issues that arose. Subsequently, we conducted some user studies to explore users' preferences regarding different interaction designs within the desktop metaverse.

This paper has two main contributions: Firstly, we discovered that despite most desktop metaverse platforms offering only a free view in the third-person mode, our results indicate a majority preference for the fixed view among users. Secondly, we introduced the innovative feature of "individual sound control," which allows users to individually

doi:10.6342/NTU202301630

mute the microphones to others and the sounds of others. Our user study revealed that this feature is necessary for most users. However, it is not widely supported in most existing desktop metaverse platforms.

doi:10.6342/NTU202301630



Chapter 2 Related Works

We explored some of the popular metaverse platforms, including *Decentraland*, *Sandbox*, *Oncyber*, *Spatial*, and *VRChat*. We compared these applications regarding text chat, voice chat, avatar movement, and perspective, as shown in Table 2.1.

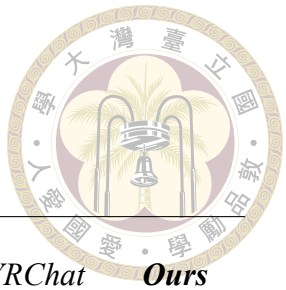
We found that all five platforms support public chat, while in *Sandbox* and *Spatial*, login is required before chatting. However, private text chat is supported only in *Decentraland*. As for voice chat, all platforms except *Sandbox* support it. Users can turn off the microphone globally in each platform but can't mute the microphones individually (i.e., choose who can hear their sound). In *Decentraland* and *VRChat*, users can mute the sounds of others individually. In addition, only in *VRChat* is remote users' sound volume adjusted by the distance to the local user.

Users can move on these platforms with keyboards, and in *Spatial*, users can also teleport with mouse clicks. Nevertheless, avatar overlapping is prevented only in *Sandbox*. In the other platforms, users can move as closed as possible, and their avatars will overlap. Furthermore, in *Decentraland* and *Sandbox*, there is only a third-person perspective, while in the rest of the platforms, both perspectives are supported. When in third-person mode, all platforms give the user a free view. A free view is that when a user drags the screen to rotate the camera, the avatar will keep its facing direction in the space. On the other

doi:10.6342/NTU202301630

hand, in a fixed view, the avatar also rotates with the camera orbiting it so that the camera is always at the back of the avatar., as shown in Figure 3.5. In subsequent sections, we will explain the free and fixed views more thoroughly.

doi:10.6342/NTU202301630



	Decentra-land	*Sandbox*	*Oncyber*	*Spatial*	*VRChat*	***Ours***
Public text chat	o	o	o	o	o	o
Private text chat	o	x	x	x	x	o
Voice chat	o	x	o	o	o	o
Mute mics to all users	o	-	o	o	o	o
Mute mics to specific users	x	-	x	x	x	o
Mute sounds of all users	o	-	x	x	o	o
Mute sounds of specific users	o	-	x	x	o	o
Adjust sound volume by distance	x	-	x	x	o	o
Keyboard control	o	o	o	o	o	o
Teleport	x	x	x	o	x	o
Avoid avatar overlapping	x	o	x	x	x	o
Perspectives (X-person)	3rd	3rd	1st, 3rd	1st, 3rd	1st, 3rd	1st, 3rd
Default perspective (X-person)	3rd	3rd	1st	3rd	1st	3rd
Free/ fixed view in 3rd-person mode	free view	free view	free view	free view	free view	both

Table 2.1: Comparison of some popular metaverse platforms and our implementation

doi:10.6342/NTU202301630



Chapter 3 System Design

3.1 System Architecture

Our implementation includes a front end (clients) and a server. The front end renders the virtual environment and interacts with the users; the server acts as a signaling server, which helps transfer the required data before the WebRTC connections are established, and handles the text message transmission.

There is a web-socket connection between each client and the server, and WebRTC peer-to-peer connections between the clients, as shown in Figure 3.1. WebRTC boasts real-time and secure audio/video streaming and offers data channels to send other data.[2] Thus, we realize voice chat with this API and synchronize the position of the users through the data channels. To keep the records, the chat messages are sent through the socket connection between the clients and the server instead of the data channels of WebRTC, where the data never pass through the server.

doi:10.6342/NTU202301630

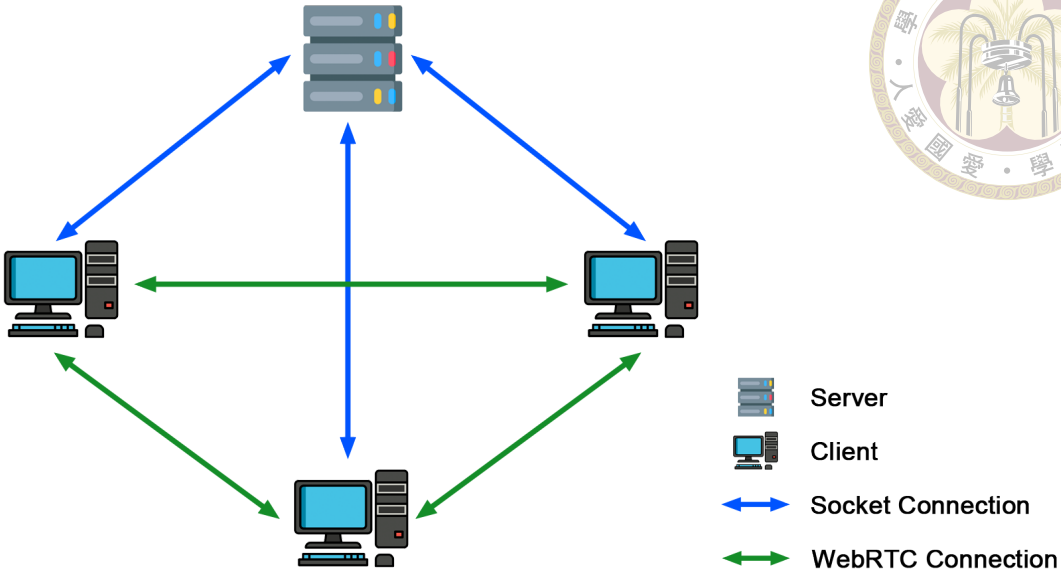
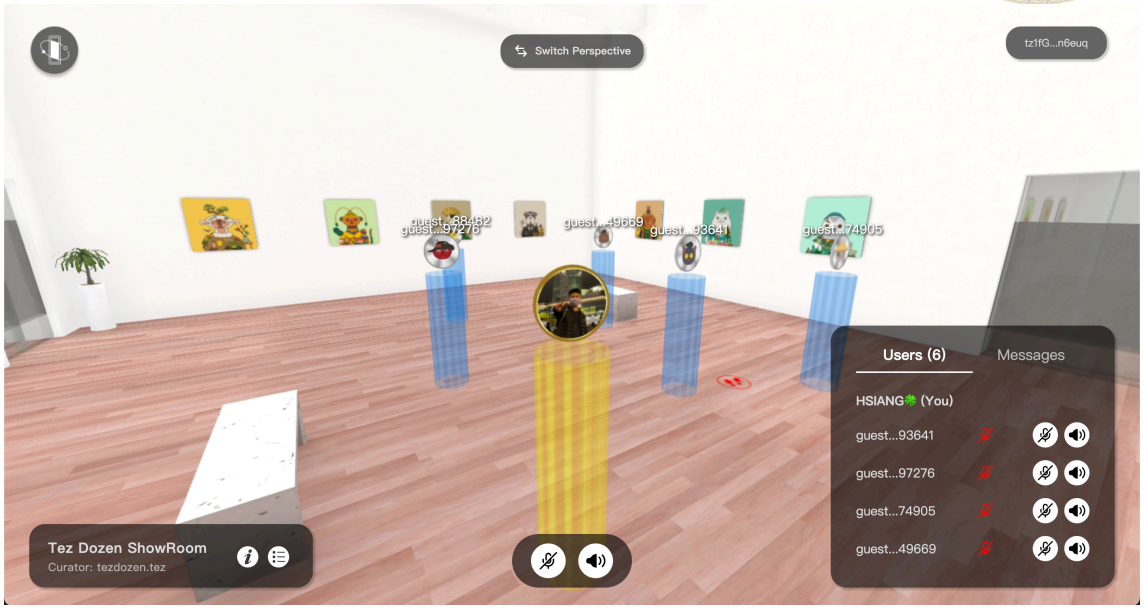


Figure 3.1: Connecting architecture of our system.

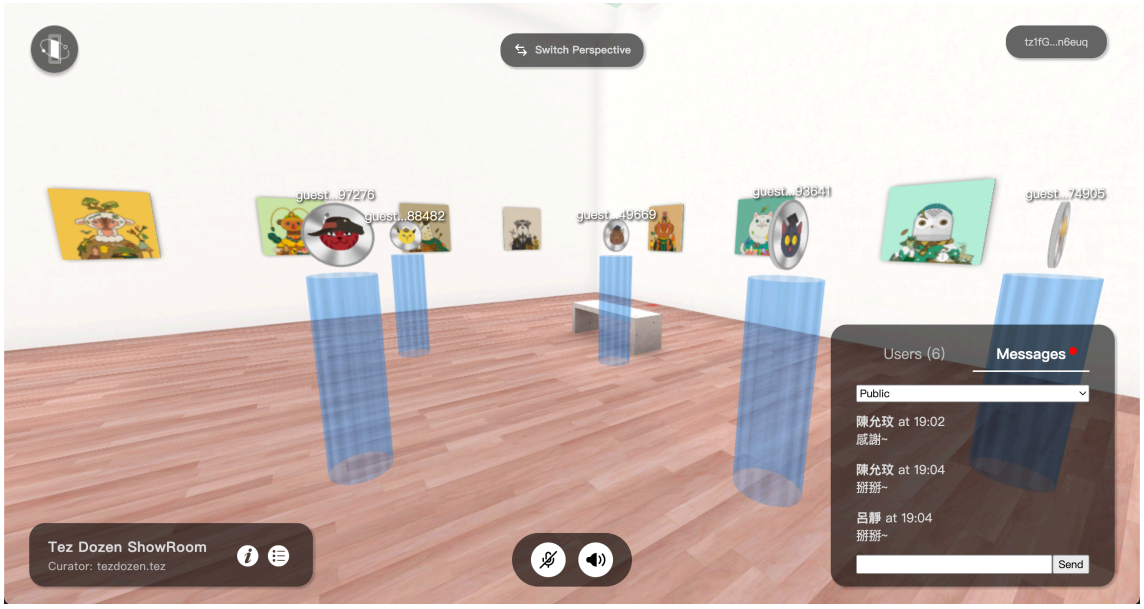
3.2 User Interface

Figure 3.2 shows what the users see when using our application. There are some avatars, panels, and buttons on the window. The avatar with a yellow cylinder belongs to the local user, while those with blue cylinders belong to remote users. The panel at the bottom right shows the online users, and the buttons at the right of each user are for individual sound control, which we will explain later. The ”Messages” tab of this panel allows users to send and read text messages, as shown in Figure 3.2b. In addition, the button at the top center with the label ”Switch Perspective” is for switching between third-person and first-person perspectives; the buttons at the bottom center are for global sound control.

doi:10.6342/NTU202301630



(a) Third-person perspective



(b) First-person perspective

Figure 3.2: User interface of our system

doi:10.6342/NTU202301630

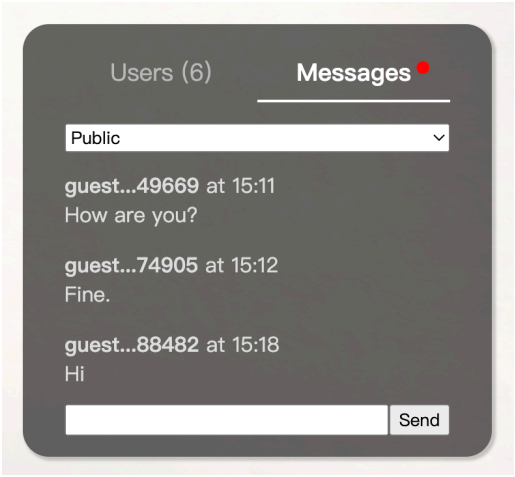
3.3 Components & Features

3.3.1 Text Chat

All the metaverse platforms we explored allow users to send text messages in a public channel, while only *Decentraland* has private channels. To give the users more flexibility, we implemented both public and private text chat. As mentioned above, the chat messages are sent through the socket to the server; then, the server keeps the chat records and transfers the messages through the socket to the receivers.

Figure 3.3a, 3.3c are the screenshots of public and private chat, respectively. This panel is at the bottom right of the window. As shown in Figure 3.3b, users can select a channel to send their messages. When there are unread messages, a red dot appears next to the ”Messages” title, and a chat bubble shows behind each channel with unread messages, as shown in Figure 3.3b. Besides, when a user sends a public message, the message will also pop on the top of his avatar for 5 seconds.

doi:10.6342/NTU202301630



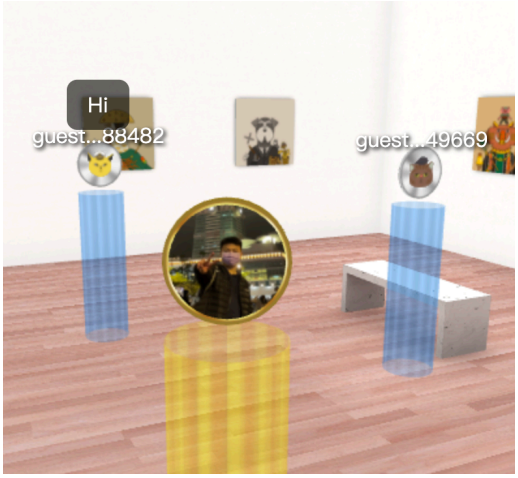
(a) Public channel



(b) Select receiver



(c) Private channel



(d) Pop message

Figure 3.3: User interface for text chat

doi:10.6342/NTU202301630

3.3.2 Voice Chat

As mentioned in Section 3.1, WebRTC offers real-time and secure channels for data transmission. Thus, we utilized WebRTC to establish peer-to-peer connections between the users. Through these connections, the audio streams are transferred to each other.

Except for *Sandbox*, all other platforms we explored support voice chat. They allow users to toggle their microphones globally. *Decentraland* and *VRChat* also enable users to mute the sounds of others individually. In our implementation, we allow users to control their microphones and the sounds of other users globally or individually. This is useful for the situation where a host or teacher introduces the virtual environment and some visitors or students want to whisper to each other, or a small group of people wants to talk without being bothered by other strangers. Furthermore, to enhance the users' immersion in the virtual environment, we also adjust the volume of other users according to their distances.

In Figure 3.2, the buttons at the bottom center of the window are for toggling the microphone and incoming sound globally. When a user mutes his microphone globally, all the other users won't hear his sound; when a user mutes the sound of others globally, he won't hear anyone in the space.

As for muting the microphones to and the sounds of others individually, a user can decide who can hear his sound and whom he wants to hear. In the rest of this paper, we will use the term "individual sound control" to refer to this feature. For example, we take the case in Figure 3.4. There are five remote users. Figure 3.4a is the panel on my window, and Figure 3.4b shows the voice data transmission between the remotes and I. The arrows represent the direction of transmission. The gray ones mean the transmission is disabled. In this situation, only users A, B, and D can hear my sound, and I will only hear the sound

doi:10.6342/NTU202301630

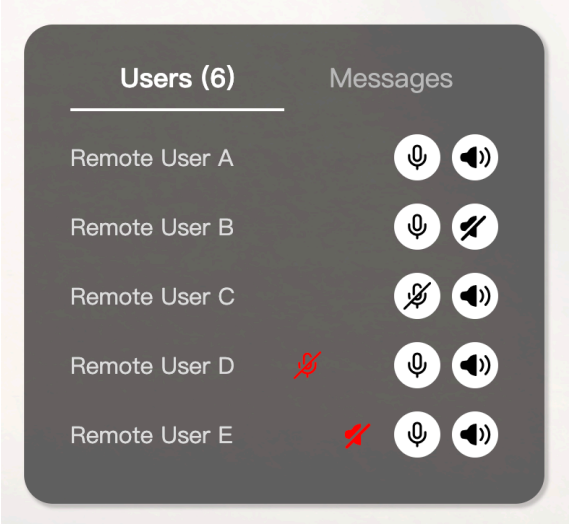
of users A, C, and E. User C can't hear my sound because I turn off my microphone to him (the microphone button is with a slash). Neither can user E because he mutes my sound. The red speaker icon with a slash behind user E means he mutes the sound of everyone or some users, including me. Similarly, I don't hear user B because I mute his sound (the speaker button is with a slash), and I don't hear user D because he turns off his microphone to me. The red microphone icon with a slash behind user D means he mutes his microphone to some users, including me.

3.3.3 Avatar

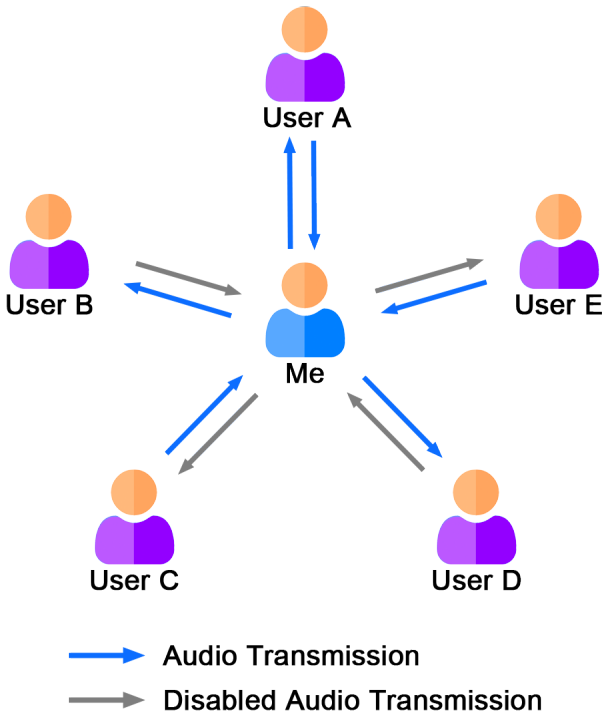
Currently, we use a coin with a picture and a semi-transparent cylinder as the avatar, as shown in Figure 3.2a. The cylinder of the local user is yellow, while the remotes' are blue. Users can move their avatars freely with a keyboard or teleport with a click. When a user moves or rotates, it will broadcast an event with the new position and rotation to other users through the WebRTC data channel. Then, the remote users can render the user's avatar in the correct position.

Sometimes, avatars may collide, and all of the platforms we explored except *Sandbox* make the avatars overlap when this happens. However, this may cause some sexual harassment issues [4]. We block a movement if another user occupies the destination to prevent this concern. Additionally, when a remote user's avatar gets too close to the local avatars during teleporting, it will become transparent.

doi:10.6342/NTU202301630



(a) User interface



(b) Voice data transmission

Figure 3.4: Example of individual sound control

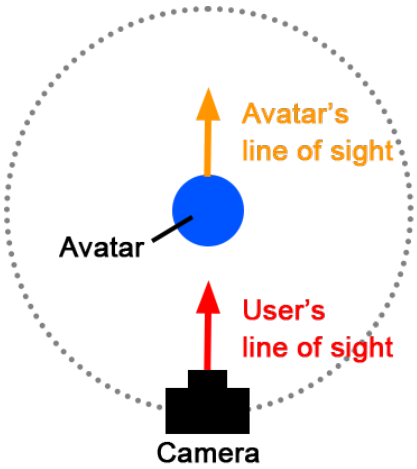
doi:10.6342/NTU202301630

3.3.4 Perspective

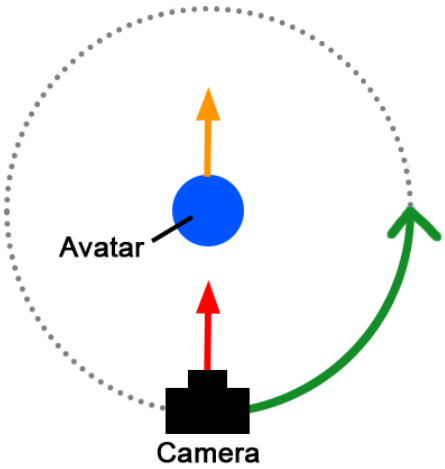
In *Decentraland* and *Sandbox*, there is only a third-person view, while others support switching between the first-person and third-person perspectives. We chose the latter to implement because some people prefer the first-person view for more immersion, while others prefer the third-person view for seeing their avatars. Figure 3.2a shows the view with the third-person perspective, the local user can see his avatar and the remote users' ; figure 3.2b is the view with the first-person perspective, the local user can only see the avatars of the remotes. Users can switch between the two perspectives with the button at the top center of the window.

In addition, we found two possible implementations of the third-person perspective: the free view and the fixed view. Figure 3.5 demonstrates the difference between a free view and a fixed view. Figure 3.5a provides a top view displaying the initial position and orientation of a user's avatar and camera. When the camera rotates, as shown in Figure 3.5b, and with the free view, the avatar remains stationary. It continues to face the previous direction, as depicted in Figure 3.5c. On the other hand, with the fixed view, both the avatar and the camera rotate together, ensuring that the camera stays positioned behind the avatar at all times, as shown in Figure 3.5d. All platforms currently only offer the free view in the third-person mode. However, we believe that the fixed view also has some advantages, such as providing a higher level of intuition. Therefore, we decided to conduct further research on the two views in the user study.

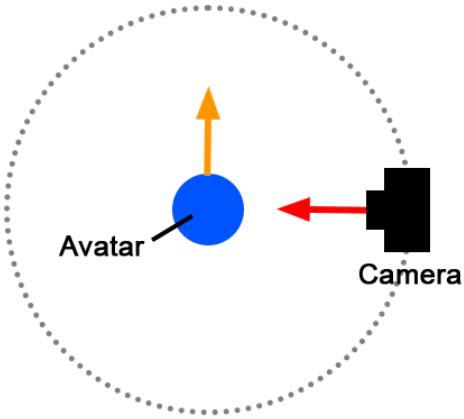
doi:10.6342/NTU202301630



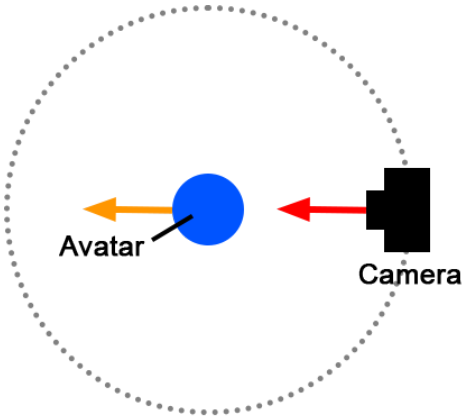
(a) At first, the user and his avatar have the same direction of line of sight



(b) Then, the camera rotates.



(c) With free view, after the camera rotates, their directions become different



(d) With fixed view, after the camera rotates, their directions are still the same

Figure 3.5: Free view vs. fixed view

doi:10.6342/NTU202301630



Chapter 4 User Study

We conducted two user studies. The first was about voice chat and perspectives, while the second was centered around voice chat.

4.1 User Study 1

4.1.1 Objectives

During the implementation, we found that some problems need more discussion. The first one is whether individual sound control is required. The second one is that we want to know whether the first-person or third-person perspective is preferable for users when exploring the desktop metaverse. The last one is whether the users prefer a free or fixed view when in the third-person mode.

Thus, we designed the user study to allow the participants to complete the following tasks:

1. Use voice chat and individual sound control.
2. Move and look around in the virtual environment with different perspectives.
3. Experience and compare the free view and fixed view.

doi:10.6342/NTU202301630

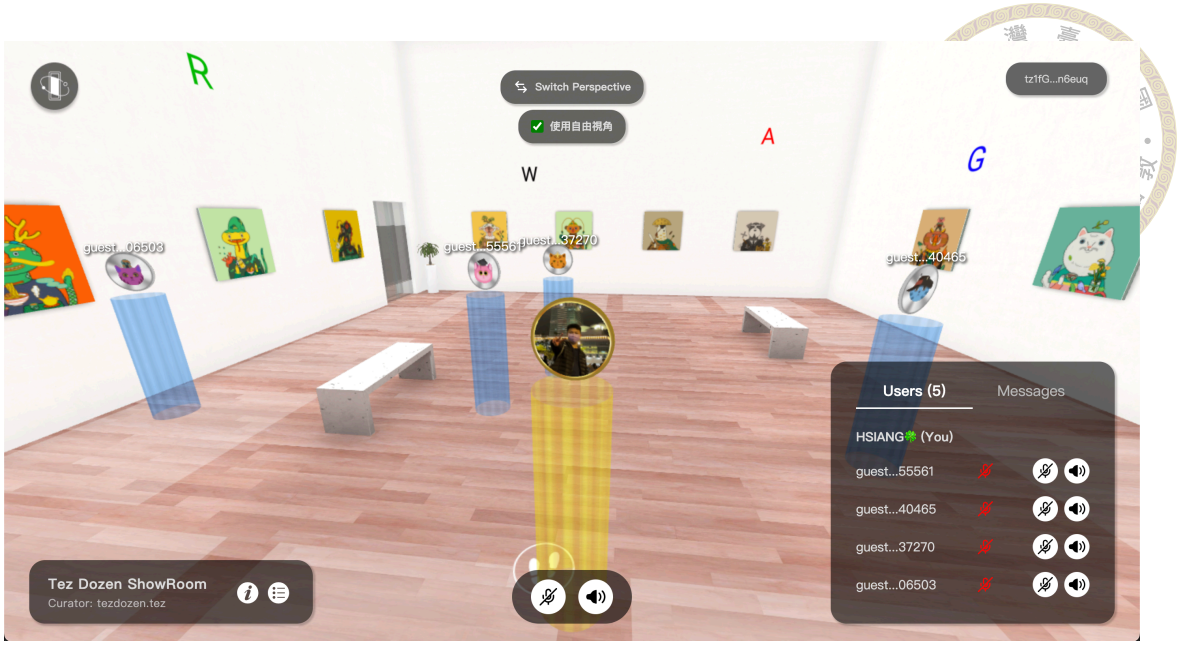


Figure 4.1: Our system with user study setups.

4.1.2 Setups

Before the user study, there were some temporary modifications in the system. First, we implemented a ”host” role. The users can’t mute the host’s sound, and the host’s sound is always with the same volume as the distance changes. By doing this, we can ensure that the participants can always hear our instructions during the test. Second, the participants must type in their names when entering the system. This is for quickly recognizing who is the owner of an avatar. Third, we added a button to switch between the free and fixed views in the third-person perspective so that the users can experience both views. Finally, we added some marks to the space. Some English letters are on the wall, and some squares are on the floor. We will talk about them later. Figure 4.1 shows the screenshot of the system with these setups.

doi:10.6342/NTU202301630

4.1.3 Process

We introduced the system to the participants at the beginning of the user study. Then, we asked them to fill in some basic information, including their names, the devices and browsers they were using, and their experience with other metaverse platforms. Then, the users started the experience. We divided the experience into four parts. In the first three parts, the participants did similar tasks but with perspectives: third-person perspective with free view, third-person perspective with fixed view, and first-person perspective.

Each of the three parts includes three tasks: For the first task, we gave the participants some questions about the pictures and letters on the wall. The second task is grouping. The participants were asked to select and move to one of the squares on the ground, and there should be 3 or 4 people on each square. The participants on the same square were in the same group and needed to complete the third task together. In the third task, we gave the participants three questions, ”What’s your favorite food?” for example. Then, they orally asked and wrote down the answers from their group members. After each part, we asked the participants some questions about their experience with the perspective used in the part.

In the final part, we asked the participants to choose their favorite six from the 20 activities we listed. Then, they talked with others, as many people as possible, finding out how many common hobbies they have with each other. After this part, some questions about individual sound control and their preferred perspective were asked.

doi:10.6342/NTU202301630

4.2 User Study 2

4.2.1 Objectives

After the first user study, we wanted to ensure whether adjusting the remotes' sound volume by distance influences the users' preference for individual sound control. Thus, we designed the user study to allow the participants to complete the following tasks:

1. Use voice chat and individual sound control.
2. Experience and compare the conditions where the remotes' sound volume is/isn't adjusted by distance.

4.2.2 Setups

In the user study, we also implemented a "host" role to prevent the participants from missing our instructions during the test. Also, the participants need to type in their names when entering the system. Moreover, we prepared two rooms. The remotes' sound volume is the same everywhere in the first room, while the volume is adjusted by distance in the second room.

4.2.3 Process

Like user study 1, we first introduced the system to the participants and asked them to fill in some basic information. Then the experience was divided into two parts: the first part occurred in the first room, and the second occurred in the second room. Similar to the fourth part of user study 1, in each part, we asked the participants to choose their favorite

doi:10.6342/NTU202301630

six from the 20 items we listed. Then, they talked with as many people as possible, finding out how many common items they selected with each other. Finally, we asked about their attitude toward individual sound control in each part and their preference between the voice chat experience in the two rooms.

doi:10.6342/NTU202301630



Chapter 5 Results

5.1 User Study 1

In User Study 1, there were a total of 21 participants, consisting of 10 males and 11 females. Among them, one participant had significant prior experience with the meta-verse platforms we investigated, five participants had some prior experience specifically with *VRChat*, while the remaining participants had limited or no experience with these platforms.

5.1.1 Voice Chat

At the end of the study, we asked the participants whether ”individually mute the sounds of others” and ”individually mute the microphones to others” are required. The rating scale ranged from 1 (strongly disagree) to 5 (strongly agree) with the statements ”I would like to mute the sounds of others individually” and ”I would like to mute my microphones to others individually.”

doi:10.6342/NTU202301630

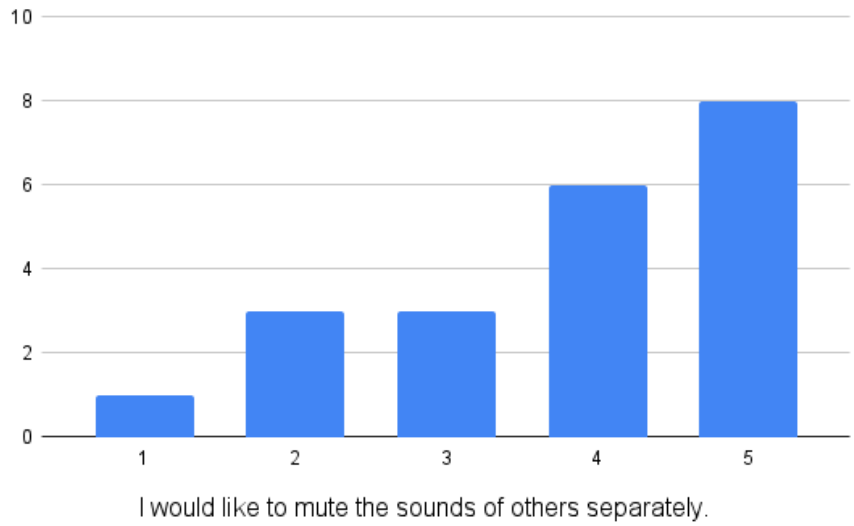


Figure 5.1: Results of ”I would like to mute the sounds of others separately.” in User Study 1 (1: strongly disagree; 5: strongly agree)

Individually mute the sounds of others

Figure 5.1 presents the results of the statement ”I would like to mute the sounds of others individually.” 67% of the participants responded positively. They mentioned that this feature helps avoid interference when conversing with specific users. However, 14% expressed no particular preference, while 19% indicated that they did not wish to use this feature in our metaverse. Some participants believed this feature is generally unnecessary because the sound of others diminishes as the distance increases. Others felt that it resembled the real world, where we could hear the sound of everyone around us. Additionally, some participants found it cumbersome to find and mute specific users.

Individually mute the microphones to others

Figure 5.2 presents the results of the statement ”I would like to mute my microphones to others individually”. 76% of the participants considered this feature necessary. Most of them mentioned privacy as the main reason. They conveyed their intention to avoid

doi:10.6342/NTU202301630

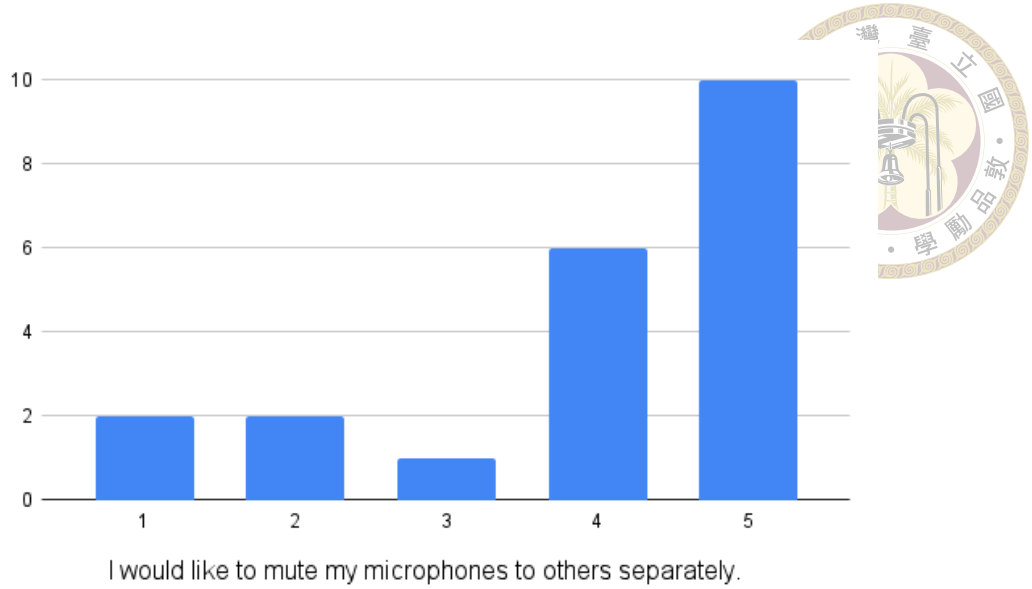


Figure 5.2: Results of "I would like to mute my microphones to others separately." in User Study 1 (1: strongly disagree; 5: strongly agree)

others hearing their conversations and prevent disturbing other users. On the other hand, some participants found this feature inconvenient. They sometimes forgot to unmute their microphones or had difficulty finding the users they wanted to mute.

Other feedback

In addition to the questions we asked, participants provided further feedback. Some recommended displaying visual indicators when a remote user is speaking to facilitate recognition of the speaker. They also proposed the functionality of clicking on other users' avatars to mute the microphone to them or the sound of them. Furthermore, participants suggested a search feature to find specific users from the user list.

doi:10.6342/NTU202301630

5.1.2 Perspectives

After each of the first three parts of the experience in the user study, we asked the participants to indicate their agreement level with the following statements on a scale from 1 (strongly disagree) to 5 (strongly agree):

1. I can feel myself wandering in the space when moving my avatar.
2. I feel I'm talking with other users face-to-face.
3. I can clearly know my position in the virtual space.
4. I like to explore the virtual space with this perspective.

First-person vs. third-person perspective

First, we focus on the comparison of results between the first-person perspective and the third-person perspective. Based on the data presented in Figure 5.3, it is evident that a significant number of participants agreed with the statement that they could feel themselves wandering in the virtual environment while moving their avatars with the two perspectives. Notably, no participant disagreed with this statement.

As shown in Figure 5.4, more participants reported feeling as though they were talking to others face-to-face when using the first-person perspective. However, a few participants responded negatively in the third-person mode with either free or fixed views. This could be attributed to the fact that we do not see ourselves while interacting with others in the real world. Conversely, with the third-person perspective, it may appear like we are watching our avatars converse with others.

doi:10.6342/NTU202301630

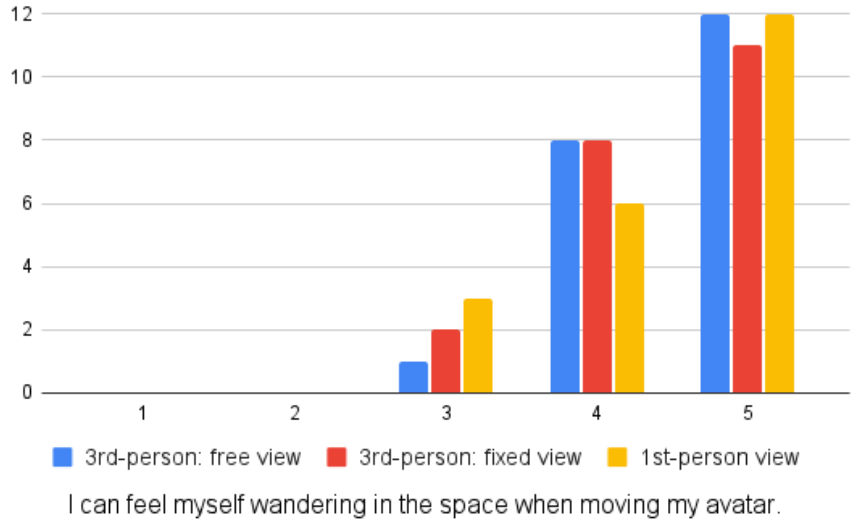


Figure 5.3: Results of ”I can feel myself wandering in the space when moving my avatar.” (1: strongly disagree; 5: strongly agree)

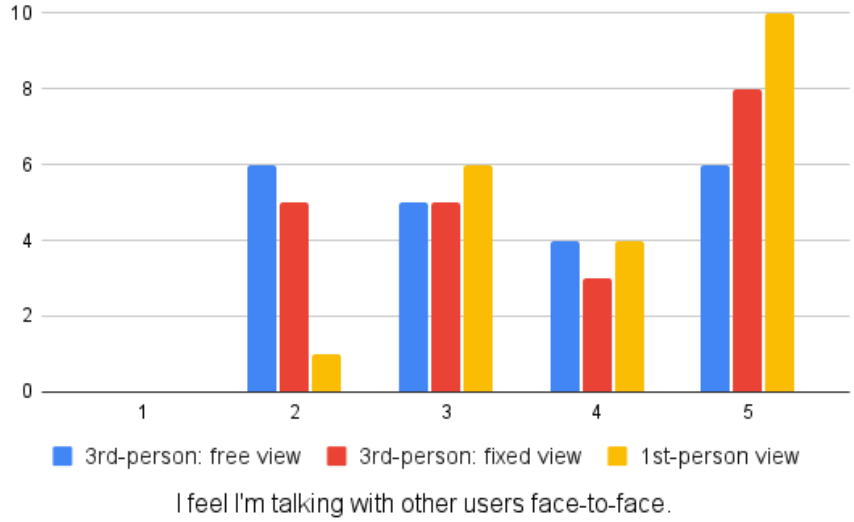


Figure 5.4: Results of ”I feel I’m talking with other users face-to-face.” (1: strongly disagree; 5: strongly agree)

doi:10.6342/NTU202301630

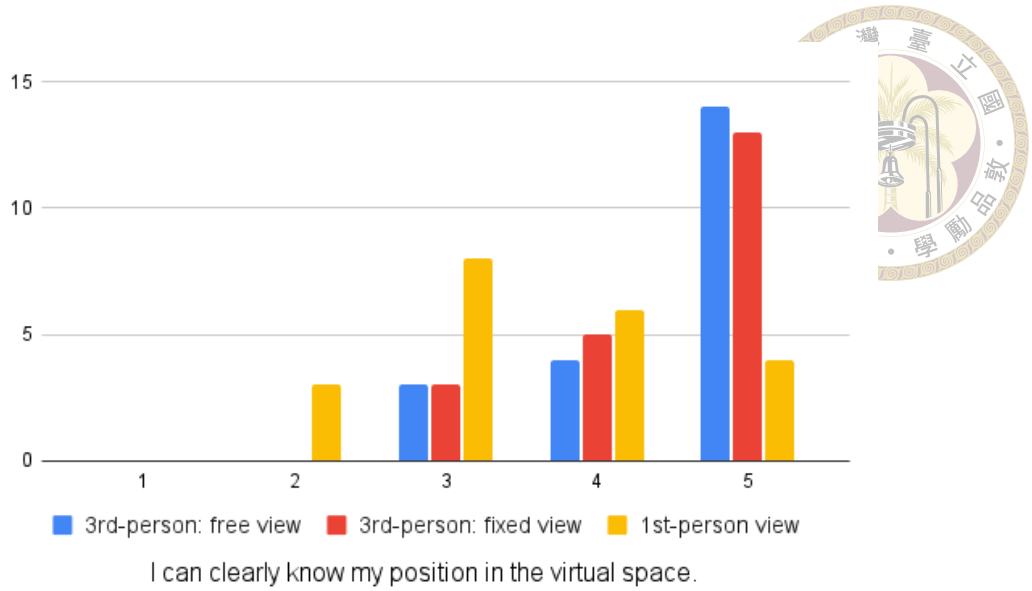


Figure 5.5: Results of "I can clearly know my position in the virtual space." (1: strongly disagree; 5: strongly agree)

The results depicted in Figure 5.5 indicate that many participants were confident about their position in the virtual space when using the third-person perspective. This can be attributed to the fact that, with the first-person perspective, users can't see their avatars. Conversely, with the third-person perspective, users can see their avatars standing in the virtual space, giving them a clearer understanding of their position.

According to the data in Figure 5.6, most participants preferred exploring the desktop metaverse using the third-person perspective. Similarly, when we asked participants to choose their favorite perspective, the results remained consistent, as demonstrated in Figure 5.7. Participants who favored the third-person perspective mentioned that it offered a broader view and reduced the chances that other avatars blocked their vision. Additionally, some participants said they experienced more vital motion sickness when using the first-person perspective. On the other hand, participants who preferred the first-person perspective argued that it gave them a heightened sense of immersion, enabling them to feel more connected and immersed within the virtual space.

doi:10.6342/NTU202301630

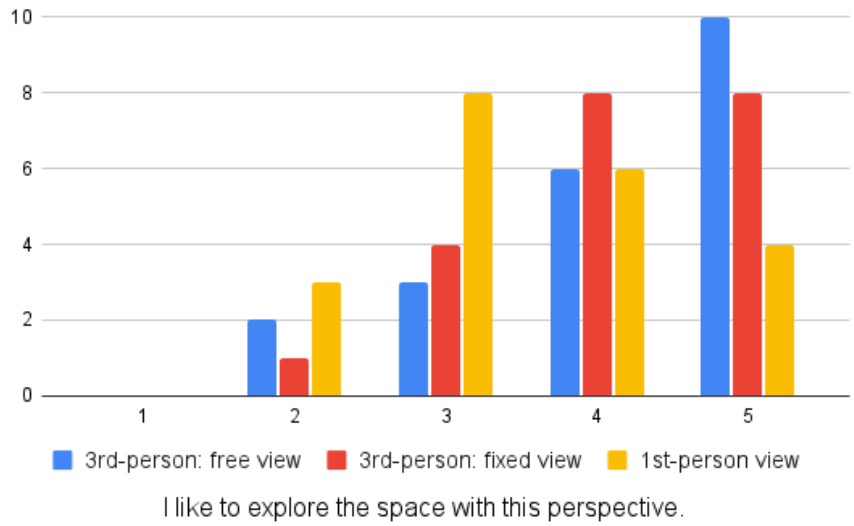


Figure 5.6: Results of "I like to explore the space with this perspective." (1: strongly disagree; 5: strongly agree)

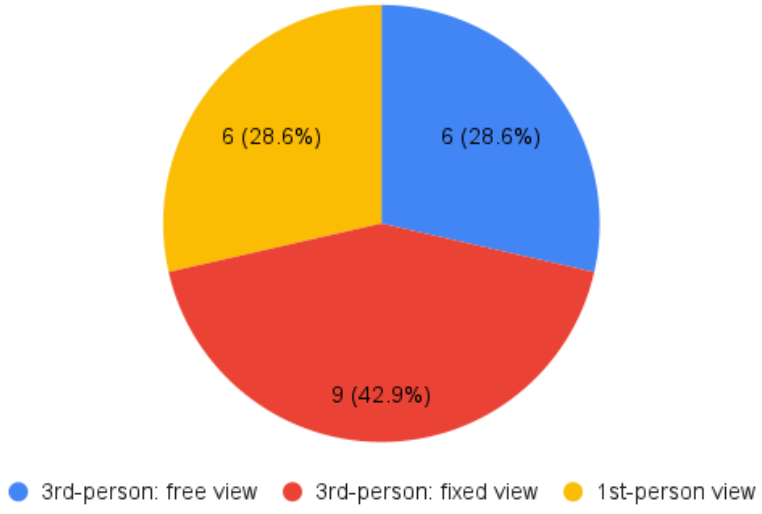


Figure 5.7: Results of "Which perspective do you like the most?"

doi:10.6342/NTU202301630

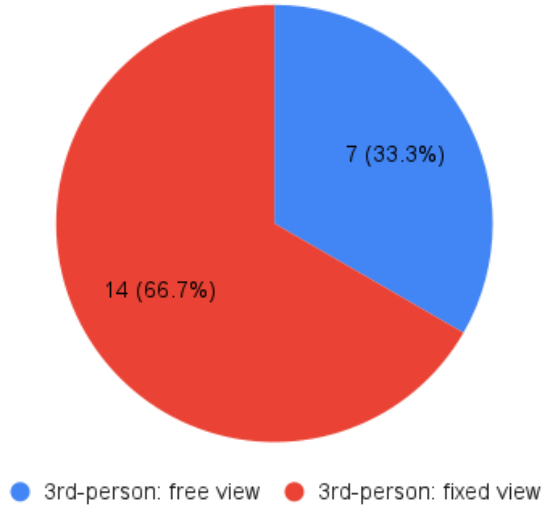


Figure 5.8: Results of "Do you like free or fixed view more?"

Free view vs. fixed view in third-person mode

Based on the data presented in Figures 5.3, 5.4, 5.5, and 5.6, participants' experiences were generally similar with the two views within the third-person perspective. When the participants were asked to choose their favorite perspective among free view, fixed view, and first-person view, or only among free view or fixed view, the majority of them opted for the fixed view, as shown in Figure 5.7 and 5.8. Participants who favored the free view said they enjoyed seeing their avatars from different angles. Additionally, the free view gave them a sense of privacy, as other users would not be aware of their specific line of sight.

On the other hand, those who preferred the fixed view expressed that it felt more intuitive. The aligned sight of the avatars allowed for a clear understanding of the exact direction of their gaze within the virtual space. Furthermore, these participants felt a stronger connection between themselves and their avatars. They experienced an enhanced sense of presence as if they were personally immersed in and observing the virtual world.

doi:10.6342/NTU202301630

5.2 User Study 2

In User Study 2, there were 19 participants, consisting of 12 males and seven females. In both user studies, the participants took the tests simultaneously. The two user studies took 40 and 20 minutes, respectively.

The experience was divided into two parts: one in a room with a constant sound volume of remotes and another in a room with sound volume adjusted by distance. At the end of each part, we asked the participants whether ”individually mute the sounds of others” and ”individually mute the microphones to others” are required. The rating scale ranged from 1 (strongly disagree) to 5 (strongly agree) with the statements ”I would like to mute the sounds of others individually” and ”I would like to mute my microphones to others,” respectively individually.

Individually mute the sounds of others

Figure 5.9 displays the results regarding the statement ”I would like to mute the sounds of others individually.” In the first room, where the sound volume of others remains constant, 68% of the participants responded positively, while 10% stated that they did not require this feature. However, in the second room, where the volume is adjusted based on the distance between the remotes and the local user, only 53% of the participants considered the feature necessary, while 21% responded negatively.

doi:10.6342/NTU202301630

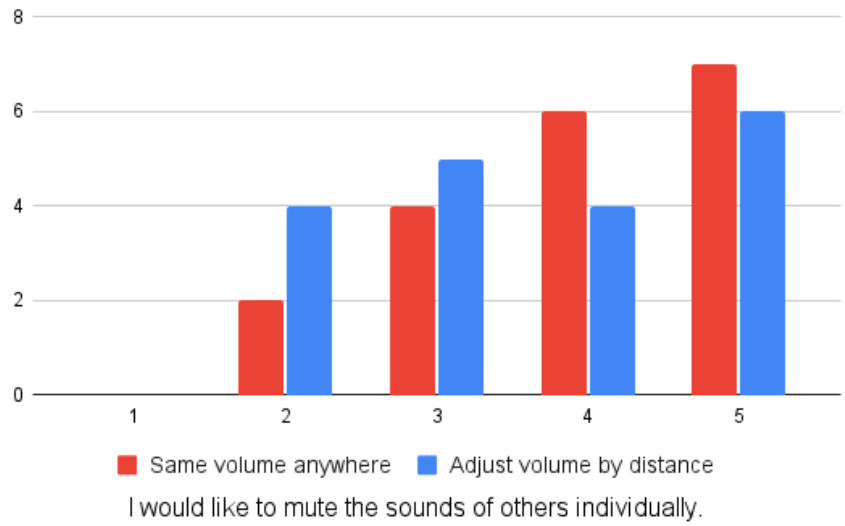


Figure 5.9: Results of "I would like to mute the sounds of others individually." in User Study 2 (1: strongly disagree; 5: strongly agree)

Individually mute the microphones to others

Figure 5.10 shows the results related to the statement "I would like to mute my microphones to others individually." In the first room, 47% of the participants deemed this feature necessary, while 26% found it unnecessary. Nevertheless, in the second room, only 42% thought the feature was required, while 37% responded negatively.

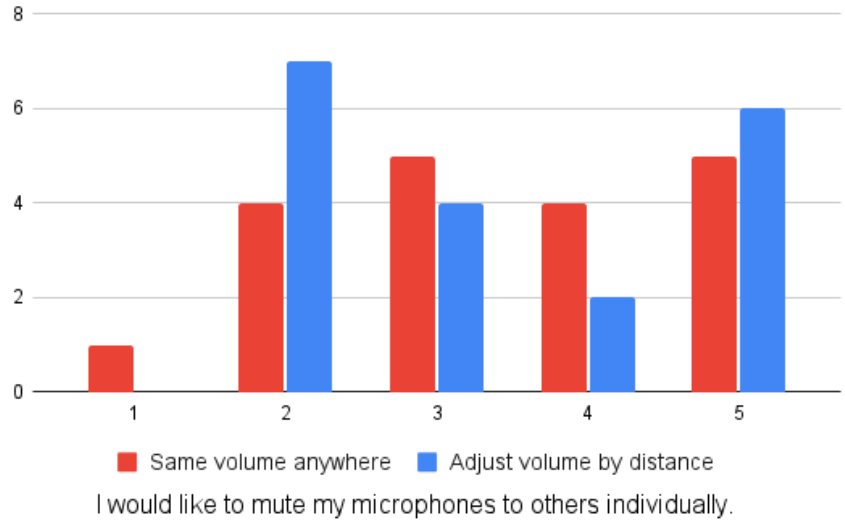


Figure 5.10: Results of "I would like to mute my microphones to others individually." in User Study 2 (1: strongly disagree; 5: strongly agree)

doi:10.6342/NTU202301630

Adjust the sound volume by distance

Finally, we asked the participants to choose their favorite room based on their voice chat experience without informing them of the differences between the rooms. As depicted in Figure 5.11, nearly all the participants selected the room where the sound volume was adjusted by distance. The only participant who chose differently mentioned that he couldn't tell any distinction between the two rooms. These results strongly indicate that adjusting sound volume by distance gives users a superior voice chat experience.

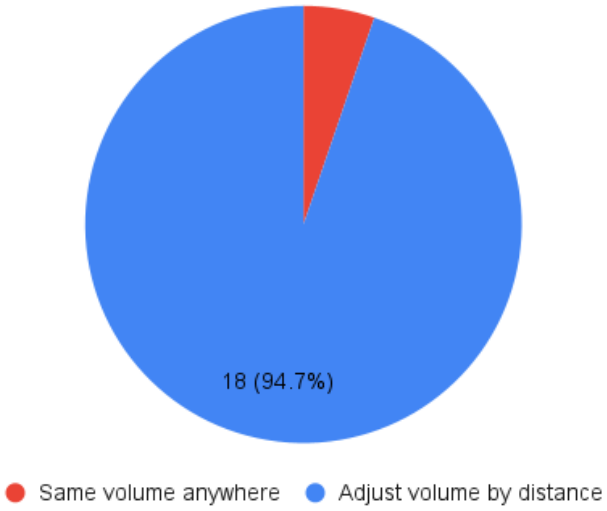


Figure 5.11: Results of "In which room did you have a better voice chat experience?"

doi:10.6342/NTU202301630



Chapter 6 Discussion

6.1 Voice Chat

Based on the results of the two user studies, it is evident that in the first user study, most participants considered ”individual sound control” necessary, while in the second study, fewer participants expressed a need for this feature. This difference might be attributed to the fact that in User Study 1, participants engaged in group chats, allowing them to interact with the same people for an extended duration. Conversely, in User Study 2, participants rapidly changed conversation partners, which could have made them feel cumbersome to use individual sound control frequently.

Moreover, the results of the second study revealed a clear preference for adjusting the sound volume of others based on distance. This setting also led to reduced demand for individual sound control. However, we believe that individual sound control remains helpful in specific scenarios, such as group exploration by friends or student group discussions.

As only a few existing metaverse platforms currently support distance-based volume adjustments and individual sound control, our findings suggest incorporating these features into future metaverse developments. However, further research is necessary to

doi:10.6342/NTU202301630

design an optimal user interface for these functions.



6.2 Perspective

In our first user study, we conducted a comparison between the first-person perspective and the third-person perspective for exploring the desktop metaverse. The results indicated a clear preference among participants for the third-person perspective, which aligns with the prevalent support for this perspective in most desktop metaverse platforms.

Furthermore, we also investigated the participants' preferences for the free and fixed views within the third-person perspective. The results revealed a preference for the fixed view. Surprisingly, we found that all the examined metaverse platforms only offered the free view option in the third-person perspective. Therefore, we recommend incorporating the fixed view into the desktop metaverse experience. It is unnecessary to replace the free view entirely. We believe there are still conditions where the free view is helpful. For example, when the users want to take photos in the space. Switching between these two views would provide users with more options and cater to their individual preferences.

6.3 System Limitation

Our system is built on WebRTC technology to manage audio streaming and some data transfer. The audio streaming relies on peer-to-peer connections between two clients, potentially leading to a considerable load due to continuous data transmission. Consequently, the maximum number of connections our system can handle is determined by the device and network affordance of the clients, impacting how many users can be accommodated

doi:10.6342/NTU202301630

simultaneously. Besides, the WebRTC API we utilize allows up to 500 peer-to-peer connections.

Our system also utilizes web socket connections between the server and clients, mainly for text messages and lightweight data to establish WebRTC connections. Although these current connections are minor and do not heavily consume bandwidth, it is crucial to consider the server's affordance as the number of users increases or when we allow large file sharing, such as images or videos, in the future.

Hence, these factors should be carefully considered while planning for scalability and capacity to ensure a seamless and enjoyable metaverse experience for all users.

doi:10.6342/NTU202301630



Chapter 7 Conclusion

In this paper, we focused on studying the interaction design for the desktop metaverse. We started by exploring the most popular metaverse platforms and compared their interaction designs regarding chat, avatar, and perspective. Building upon this research, we developed our metaverse platform with text chat, voice chat, avatars, and camera perspective features. During the development process, we identified specific points requiring further investigation and conducted two user studies to address them.

In the first user study, we explored users' preferences for "individual sound control," their preference for first-person or third-person perspectives, and their preference for free or fixed views within the third-person mode. The study involved 21 participants, and the results indicated that most participants found "individual sound control" necessary and preferred the third-person perspective, particularly with the fixed view.

In the second user study, we emphasized voice chat. We examined users' need for individual sound control when the sound volume of remote users was adjusted based on the distance to the local user. The study involved 19 participants, and the results showed that fewer users required individual sound control under such conditions, mainly when frequent changes in conversation partners occurred.

Based on our findings, we propose several suggestions for enhancing the interaction

doi:10.6342/NTU202301630

design for the desktop metaverse. First, we recommend adjusting the sound volume based on the distance to the local user for each remote user. Second, it would be beneficial to provide individual sound control, allowing users to mute their microphones to others and the sound of others individually. Last, including the fixed view as an option within the third-person mode would offer users greater flexibility. By incorporating these suggestions into the interaction design, we can optimize the desktop metaverse's usability and provide users with a more enjoyable and immersive experience.

doi:10.6342/NTU202301630



Chapter 8 Future Work

Introduce Video Chat

We plan to enable video streaming through WebRTC to introduce video chat functionality in our metaverse platform for future work. While popular applications like Google Meet, Zoom, and Teams provide real-time audio and video streaming for online meetings, they lack a 3D virtual world experience and the ability to have private conversations with selected users. In our metaverse, we aim to design an online meeting room where users can conduct meetings as if they are physically present in the same space. Additionally, users can whisper or have private conversations with specific individuals, enhancing the immersive and interactive nature of the metaverse experience. This feature will create a unique and engaging virtual meeting environment, offering a novel online collaboration and communication approach.

Human-Like Avatar

We can enhance the user experience by significantly improving the avatar feature. One key aspect is to replace the current avatar with a more sophisticated and human-like 3D model. This upgrade aims to increase user immersion and identification with

doi:10.6342/NTU202301630

their avatars. We will also conduct further research to study the impact of human-like avatars on user immersion and the social distance issue in the metaverse. Additionally, we will explore integrating computer vision techniques into our system to enhance user experience. For instance, we can use webcams to detect users' faces and enable real-time avatar facial expression changes or detect gestures to control the avatars.

Special Roles

Introducing special roles with specific permissions is another aspect of our future work. For instance, we plan to implement a "host" role, allowing designated users to act as guides or teachers. Hosts can mute other users' microphones, ensuring uninterrupted communication during introductions or instructional sessions. This feature will enhance the overall user experience and foster smooth interactions, creating a more enjoyable environment for all participants.

doi:10.6342/NTU202301630



References

[1] akaVerse Inc. akaVerse, 2022.

[2] N. Blum, S. Lachapelle, and H. Alvestrand. WebRTC. Communications of the ACM, 64(8):50–54, July 2021.

[3] Decentraland Inc. Decentraland, 2020.

[4] A. Diaz. Disturbing reports of sexual assaults in the metaverse: “it’s a free show”. New York Post.

[5] Y. Feng, D. C. Duives, and S. P. Hoogendoorn. Wayfinding behaviour in a multi-level building: A comparative study of HMD VR and desktop VR. Advanced Engineering Informatics, 51:101475, Jan. 2022.

[6] E. A.-L. Lee and K. W. Wong. Learning with desktop virtual reality: Low spatial ability learners are more positively affected. Computers &amp Education, 79:49–58, Oct. 2014.

[7] E. A.-L. Lee, K. W. Wong, and C. C. Fung. How does desktop virtual reality enhance learning outcomes? a structural equation modeling approach. Computers &amp Education, 55(4):1424–1442, Dec. 2010.

[8] Meta Inc. What is the metaverse?, 2023.

doi:10.6342/NTU202301630

[9] Oncyber Inc. Oncyber, 2021.

[10] Spatial Systems Inc. Spatial, 2018.

[11] P. Srivastava, A. Rimzhim, P. Vijay, S. Singh, and S. Chandra. Desktop VR is better than non-ambulatory HMD VR for spatial learning. Frontiers in Robotics and AI, 6, July 2019.

[12] The Sandbox Inc. The Sandbox, 2019.

[13] Verified Market Research. Metaverse Market Size, Share Growth | Industry Report., Mar. 2023.

[14] VRChat Inc. VRChat, 2017.

[15] F. Xue, R. Guo, S. Yao, L. Wang, and K.-L. Ma. From artifacts to outcomes: Comparison of HMD VR, desktop, and slides lectures for food microbiology laboratory instruction. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. ACM, Apr. 2023.

doi:10.6342/NTU202301630