Taylor & Francis
Taylor & Francis Group

International Journal of Human–Computer Interaction

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/hihc20

Examining the Use of *DanMu* for Crowdsourcing Control in Virtual Gatherings

Yue Li, Teng Ma, Ziming Li & Hai-Ning Liang

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2024.2375700

Examining the Use of *DanMu* for Crowdsourcing Control in Virtual Gatherings

Yue Li[a], Teng Ma[a], Ziming Li[a], and Hai-Ning Liang[b]

[a]School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou, China; [b]Computational Media and Arts Thrust, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China

ABSTRACT

The advent of web-based interactive technologies has opened up new possibilities for virtual gatherings in 3D environments. Live-streaming, in particular, has gained increasing attention due to its effectiveness in engaging a large number of users in collective online activities. With an emphasis on audience participation, live-streaming shares common characteristics of the outlook of the metaverse and is driving new waves of interaction in virtual gatherings, such as engaging users through crowdsourcing control. However, this type of social interaction has not been examined in the Asian context, and it lacks systematic investigation of user experience with different crowdsourcing control methods. In this paper, we present a novel crowdsourcing control method based on DanMu, the subtitle system of Bilibili, one of the most successful and prevalent live-streaming platforms in Asia. We organized virtual gatherings by live-streaming a Minecraft virtual campus and examined the use of DanMu for crowdsourcing control. Our first study investigated the influence of three crowdsourcing control methods (First Come First Served, Vote, and Super Command) on collective navigation task efficiency and user experience. These influences were further discussed with user activeness and group sizes in a follow-up study. The results showed that Super Command, a representative mode on top of the democratic voting mechanism, offers better user experiences and social richness in large groups. Participants also rated its usability higher in small groups. Besides, virtual gathering in small groups allows greater pragmatic quality, usability, and a sense of agency than in large groups. Our work provides design guidelines for developers and HCI practitioners to develop crowdsourcing control methods and improve novel virtual gathering experiences in virtual worlds and the future metaverse.

KEYWORDS

Multiplayer single-character control; user experience; virtual gathering; live-steaming; audience participation

1. Introduction

The metaverse, a term popularized by science fiction and now gaining substantial traction in technology discourse, refers to a collective virtual shared space, merging the physical and digital worlds (Lee et al., 2021). In recent years, advancements in virtual reality (VR) and augmented reality (AR) technologies have propelled the concept of the metaverse into tangible realms, allowing them to be together in one virtual environment and interact with each other and the virtual objects (Chen, Liang, et al., 2021; Chen, Liu, et al., 2021; Liang et al., 2019; Xu et al., 2024). Users can virtually meet together in a computer simulated environment while being physically apart. Virtual gatherings within the metaverse represent a dynamic shift in the way people interact, communicate, and collaborate remotely, supporting a large number of users to have collective activities in 3D virtual worlds and generate data that influence the virtual worlds (Getzlaf et al., 2012; He et al., 2023). While existing video conferencing systems such as *Zoom* support virtual gathering activities in people's daily work and life, future virtual gatherings are likely to take place in 3D virtual environments, showing digital twin environments that closely resemble the real environments or completely fictional worlds (Shiau & Huang, 2023). Such 3D virtual environments will support users' spatial awareness and the sense of *being there* in the environments. Some existing platforms support virtual gatherings in 3D virtual worlds, such as *Spatial.io* and *Mozilla Hubs*, but they could host a limited number of users (eg, up to fifty). There is a gap between the visual fidelity and the capacity to host such virtual gathering activities. However, one notable scenario, despite some limitations in the degree of control, supports a large number of users to gather in high-fidelity 3D virtual worlds: live-streaming.

Live-streaming refers to the real-time transmission of video or audio content over the internet. In a live-stream, the content is broadcasted as it is happening, allowing thousands of users or more to watch and, in many cases, interact with the stream in real-time (He et al., 2023). It supports large-scale concurrent users and engages them in active participation, of which the social characteristics are similar to the metaverse. Notably, Bilibili is one of the most popular live-streaming platforms in China, with over 300 million monthly active users (Statista, 2023). It is known for its unique *DanMu* (bullet comment) feature, showing real-time

CONTACT Hai-Ning Liang haininglaing@hkust-gz.edu.cn Computational Media and Arts Thrust, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China
This article has been republished with minor changes. These changes do not impact the academic content of the article.

comments posted by users that scroll across the screen like a bullet while watching a video or a live-stream (Wang et al., 2019; Zhang et al., 2023). This creates a dynamic and interactive viewing experience, allowing users to share reactions and thoughts in real-time. Previous works showed that audience participation contributes to user attitudes towards live-streams and enhances their perceived value in participating in social interactions (Chen & Lin, 2018; Striner et al., 2021; Zhang et al., 2023). Existing approaches to engaging audience participation in live-streaming include likes, comments, and gifts, among others, most of which add an additional layer of information on top of the streamed content (Li & Guo, 2021). However, these approaches do not allow users to directly affect the streamed content. User control over live-streaming content introduces a participatory element to the gathering experience, offering a more engaging form of interaction that aligns with the evolving expectations of modern digital audiences.

Crowdsourcing control is one of the most engaging approaches to engaging users' active participation in virtual gatherings. It allows a large group of individuals, often referred to as the "crowd", to collectively contribute to decision-making, problem-solving, or control of a system (Lesniak & Maistro, 2022). One typical example of crowdsourcing control is multiplayer single-character control (MSCC). As the name suggests, it allows multiple users or players to collectively control a single character in a game. Ever since the phenomenal experiment of *Twitch Plays Pokémon (TPP)*, which holds a Guinness world record of gathering over a million players on a single-player online videogame, MSCC has been applied to different types of games and contents (Ramirez et al., 2014; Seering et al., 2017). Players are often allowed to adopt two methods of control: anarchy and democracy. In anarchy mode, every viewer's input is registered based on a *"first come first served"* mechanism. This results in a chaotic and often unpredictable gameplay experience, as numerous commands are simultaneously entered, leading to diverse and sometimes conflicting actions. In democracy mode, the system tallies the inputs sent by players over a short period (eg, five seconds) and executes the most popular command. This mode introduces a *voting* mechanism and allows for more coordinated and strategic gameplay. Recent works also explored variations of the two types of crowdsourcing control (Lesniak & Maistro, 2022). MSCC provided participants with a unique and engaging way to participate in the game and obtain a sense of achievement (Mallory, 2014). While MSCC has been studied in chatbox environments within English-speaking countries, we did not find any related work investigating its use in the Asian context. In particular, the use of DanMu as a control method has not been examined in previous works. Unlike the English Latin alphabet, Asian languages such as Chinese use logographic symbols representing words or morphemes. Thus, there are potential socio-cultural differences in the user experience.

Previous works showed that MSCC faces challenges in the design of interactions and user experience, such as the sense of agency (Striner et al., 2021). Users engaged in MSCC within virtual worlds often seek to conduct meaningful actions as a whole while each user has individual characteristics (Lesniak & Maistro, 2022). This indicates the need to understand explicit user experience, addressing both the pragmatic and hedonic aspects. Existing studies about crowdsourcing control have tried to design more effective ways to aggregate participants' commands in collective behaviors (Lesniak & Maistro, 2022). However, a user experience perspective is missing. In particular, while anarchy and democracy modes emerged as two notable methods of MSCC, and previous work implied their relations with task efficiency and unpredictability of games (Lytle et al., 2020; Mallory, 2014), there was no systematic study carried out to show their effects on usability and user experience. Motivated by this research gap, we ask **RQ1:** *How do multiplayer single-character control methods influence task efficiency and user experience in virtual gatherings?*

Aside from the control methods, users' collective behaviors and their experiences are largely affected by the context where the virtual gathering activity happens. Existing studies have shed light on the effects of peer influence on the varied outcomes of group performance and perceptions (Carillo & Okoli, 2011; Ford et al., 2017). For example, a large group comes with frequent communication and also faces challenges of high density of information, which may negatively influence the outcome of group activities (Carillo & Okoli, 2011). Similarly, active users in a group could stimulate the herding effect among the audience to participate in the activity but may also impose negative feelings due to the competition for attention (Fan et al., 2018). Thus, we propose **RQ2:** *How do user activeness and group size influence task efficiency and user experience in virtual gatherings?*

In this paper, we examine the use of DanMu for crowdsourcing control in virtual gatherings. Based on the two main control methods (anarchy: First Come First Served; democracy: Vote) identified from existing works (Pascal et al., 2017; Ramirez et al., 2014), we propose Super Command, a variation of the Vote method that adds a representative power on top of it. We use Minecraft as our experimental platform. It has a large and diverse user base, including players of different ages and backgrounds. In addition, it provides an open-ended and customizable virtual environment, offering researchers the flexibility to create diverse scenarios for experimentation. Its modding community and open API also make it adaptable for research purposes. Thus, we created a virtual world in Minecraft and conducted two user studies. Our first study revealed that comparing the three methods of control, Super Command supported the greatest pragmatic and hedonic experience, usability, perceived performance, body ownership, sense of agency, and social richness. Based on the number of DanMu they sent, we invited ten active and ten inactive participants to rejoin and participate in a small group gathering. The results showed that participants in the active group perceived greater hedonic experience and social richness than those in the inactive group. Participants also perceived greater pragmatic quality, usability, and a sense of agency in small groups than in large groups. Based on the study findings, we offer design recommendations for developers and HCI practitioners. These guidelines aim to enhance the user experience of crowdsourcing control in virtual gatherings.

The contributions of our work are three-fold. First, we propose a novel web-based approach with text-based command line inputs (ie, DanMu) that allows crowd users to interact with 3D virtual worlds. Second, two empirical studies are presented, investigating the effects of control methods, user activeness, and group size on task efficiency and user experience in virtual gatherings. Third, informed by the subjective (self-reported questionnaires) and objective (user activities) data analysis results, a series of design guidelines were summarized, which provide valuable insights for future research and practices on virtual gathering activities towards the metaverse.

2. Related work

2.1. Virtual gatherings in game live-streaming

Over the past few decades, the landscape of social interaction has undergone a transformative shift with the emergence of virtual gatherings (Getzlaf et al., 2012; Lu et al., 2018). These gatherings represent a paradigmatic evolution in how individuals and groups assemble, communicate, and collaborate, enabled by advancements in digital communication technologies. The concept of virtual gathering encapsulates the notion of people coming together in online spaces to engage in various forms of social, professional, or communal interactions, transcending geographical limitations (Pires & Simon, 2015).

Specifically, the emergence of game live-streaming, serving as a form of virtual gathering, has received increasing attention in the field of human-computer interaction (Glickman et al., 2018; Seering et al., 2017). Its abilities in connecting individuals and creating virtual spaces for social activities have been widely recognized. For instance, the #PlayApartTogether, a promotional campaign that involves bringing special events, activities, and inspirations of games on live-streaming, was developed as a coping strategy for people to socialize during special periods such as COVID-19 (King et al., 2020). Besides, mainstream sports such as football and darts have likewise adapted live-streaming platforms to virtually engage a broad range of viewers together (Liu et al., 2022; Striner et al., 2021). Covering a wide range of themes and fields, online live-streaming has received significant traffic for game-related content (Lytle et al., 2020; Poirier-Poulin, 2020). For example, *Twitch* has gained widespread popularity in online streaming communities for a range of game content. Individual streamers use Twitch to broadcast their gameplay and comment on their experience with popular games such as *Grant Theft Auto V* and *League of Legends*. These two games have gained 25 million and 19 million hours of view per week, and their average weekly viewers have reached 150k and 110k, respectively (TwitchTracker, 2024).

Within live-streaming platforms, viewers typically react to the content, streamers, and peers through text-based communication channels (He et al., 2023; Li & Guo, 2021), which shifts the user experience of the medium from spectatorship, like traditional television experience, to online social interactions. For instance, the contents posted by viewers, including DanMu, emojis, and gifts, are displayed on top of the video content to show the real-time reactions from the audience, forming vigorous conversations in the shared space (Deng et al., 2015). Participants see this information as lively and interactive channels through which they could communicate and socialize with both streamers and other viewers (Li & Guo, 2021; Zhou et al., 2019).

With the rise of the crowdsourcing control system, where players make collective decisions regarding the gameplay (Lesniak & Maistro, 2022), social interactions include not only the conversation situated around streamed content but also the immediacy of control over the shared gathering spaces (Haimson & Tang, 2017). Specifically, Striner et al. (2019) defined audience interactivity as a range of experiences and roles that allow audiences to participate in or interact with. The authors presented a spectrum of audience interactivity in entertainment, showing that passive observations in traditional watching experience sit at the bottom of the spectrum, whereas crowdsourcing controls “augment overall experience” and allow users to “become a performer”, thus providing a much higher level of interactivity in the spectrum.

Along with the increasing adoption of crowdsourcing controls in game live-streaming, efforts have been observed in providing effective and usable systems to organize the crowdsourcing controls from the audience (Glickman et al., 2018; Lesniak & Maistro, 2022; Pascal et al., 2017). Nonetheless, a more predictable or reliable system does not always lead to a favored user experience (Poirier-Poulin, 2020; Ramirez et al., 2014). Back to the earlier-mentioned example of *Twitch Plays Pokémon*, although the democracy mode is more predictable and stable in collecting the majority controls from the audience, the game was mostly played in anarchy mode, which leads to a higher level of chaos by random audience choices (Pascal et al., 2017; Ramirez et al., 2014). This interesting phenomenon indicates that a well-engineered system in organizing reasonable controls may not always be the optimal solution to maintain a satisfying user experience. Especially in the context of game live-streaming, the social interactions are afforded by a wide range of discourse found in the massive concurrent viewership (Ford et al., 2017), including seemingly incomprehensible or meaningless ones such as memeing and spamming. The social needs of the audience may also extend from meaningful conversations to showing off and forming “crowdspeak” that could be heard. Thus, the difference in crowdsourcing control methods needs to be examined to provide useful design guidelines for virtual gatherings.

2.2. Crowdsourcing control methods and applications

Among various crowdsourcing control systems, the one where multiple players control a single character, termed as multiplayer single-character control, could reflect the common pattern of collective behaviors of viewers (Brandis & Bozkurt, 2021; Ramirez et al., 2014). *Twitch Plays Pokémon*, a typical example of MMSC, allowed players to simultaneously control a Pokémon game character without any kind of central authority. Due to the novelty of this game-play mechanism, it achieved great success in terms of the number of players and community development. The same approach continued successfully with other Pokemon games (Brandis & Bozkurt, 2021; Lessel et al., 2017). Users also utilized crowdsourcing controls in non-gaming areas, such as installing a Linux operating system (TwitchInstallsArchLinux, 2024). Still, most applications of MMSC were in games, and more *TPP*-like channels with crowdsourcing controls appeared. Notably, the MSCC method was also applied to Dark Souls (TwitchPlaysDark, 2024), a challenging action role-playing game that took 43 days and endless trials to complete. However, the method for crowdsourcing control had to be modified between anarchy and democracy modes to support appropriate actions in *Dark Souls*, which further leads to a wide discussion on how to coordinate and organize chaotic collective behaviors in MSCC.

Selecting and performing a specific command out of massive controls from concurrent viewership is a key factor in the success of MSCC methods (Pascal et al., 2017; Ramirez et al., 2014). In the earliest version of *TPP*, the interaction was based on a simple *first come first served* (FCFS) or *anarchy* selection mechanism: the first command that appears within a collection window will be performed (Lasecki et al., 2011). Yet, this approach faced a great challenge when a movement-based puzzle was presented in the game (Brandis & Bozkurt, 2021). The players were stuck for 24 h, so the system was changed to perform the command that got voted the most in every five seconds, which was later termed as the majority *vote* system (Ramirez et al., 2014). However, such a *democracy* system did not receive the expected praise from the crowd. Instead, participants started to protest by sending `start9`, the advanced command that will open and close the menu repeatedly, to impede any movement voted. Consequently, the event organizer had to revert the system back and allow the option to switch between the *democracy* and *anarchy* modes. Thus, we propose **H1**: *Democracy MSCC supports higher task efficiency (**H1a**) and anarchy MSCC allows greater user experience (**H1b**) in virtual gatherings.*

Previous works indicated that user experience with crowdsourcing control methods may vary with different user characteristics. For example, Brandis and Bozkurt (2021) found a correlation between user activeness and preference for control modes: highly active users favor anarchy mode, while less active users tend to prefer democracy mode. Besides, the number of viewers in live-streaming was found to shape the streaming content: streamers are more willing to do routine activities when fewer people are watching (Poirier-Poulin, 2020). In addition, viewers often prefer smaller channels if they are motivated by social engagement with other users in live-streaming (Hilvert-Bruce et al., 2018). These findings indicate that the group size, as well as the willingness to participate in collective behaviors, will affect the user experience in virtual gatherings. Based on the implications drawn from previous works, we propose **H2**: *Active users (**H2a**) and small groups (**H2b**) have greater task efficiency and user experience in virtual gatherings.*

Although existing studies have provided some insights on the relationship between MSCC methods and participant's experience and behaviors (Aleta & Moreno, 2019; Brandis & Bozkurt, 2021; Lessel et al., 2017), few of them have empirically examined the influence of different MSCC methods on the perception and collective dynamic of audience. In the next section, we discuss the necessary user experience measures of crowdsourcing control in virtual gatherings.

2.3. User experience of crowdsourcing control in virtual gatherings

As depicted by Benford et al. (2001), collaborative experience in virtual environments faces several challenges, such as scalability, interest management, and distributed architecture. Emerging virtual gathering activities such as live-streaming effectively facilitate simultaneous communications of a large number of users and integrate interactions from various ubiquitous devices, such as mobile phones and tablets (Hu et al., 2017). However, the display sizes and visual details could potentially dampen the sense of spatial presence (Slater et al., 1994), which is an important dimension of user experience in the virtual environment. In addition, as the degree of audience participation increases to allow the control of streamed content, it faces challenges in aligning collective behaviors with a consideration of human factors and user experiences (Li et al., 2020), especially in the context of crowdsourcing control.

Control is an important design factor in HCI (Sedig & Parsons, 2016). Understanding user experience associated with control methods demands the understanding of the low-level implicit sense of control and the high-level explicit experience of control (Liang et al., 2010; Liang & Sedig, 2010; Synofzik et al., 2008). Implicit sense of control often requires objective measures and participants' focused attention in repetitive trials within controlled settings, making it difficult to measure in remote studies with multiple users. On the other hand, explicit user experience of control methods can be measured using questionnaires. Bergström et al. (2022) formulated a questionnaire from various established questionnaires to measure the explicit user experience of control. Specifically, they included the AttrakDiff to measure pragmatic and hedonic user experience with technology use (Hassenzahl, 2004). In addition, the UMUX-LITE questionnaire was used to measure usability (Lewis, 2018), and the NASA Task Load Index was used to measure workload when using the control methods (Hart & Staveland, 1988). While these are generic measures that have been widely adopted to understand user experience with technologies, body ownership and the sense of agency were two important measures that are closely related to user experience with control methods (Longo & Haggard, 2009). These reflect users' immediate and pre-reflective experience of control.

Different from control methods for individual uses, crowdsourcing control engages multiple users to participate in and affect the virtual gathering experience. For instance, *Choice Chamber* allows audiences to influence in-game challenges by adding items and buffs through chat channels (StudioBean, 2016). The capacity to alter experiences has empowered audiences beyond the traditional role of passive spectators to become active participants (Brandis & Bozkurt, 2021; Striner et al., 2021). Cerratto-Pargman et al. (2014) further explained that these interactions are integral elements of a social play experience. To evaluate user experience within a social environment, the sense of social richness emerges as an important dimension of user experience (Lombard et al., 2011). This indicates how much a medium is seen as sociable, warm, sensitive, personal, or intimate when utilized for interacting with others. While spatial presence is the sense of "being there" in the virtual environment, social richness denotes the communication perspective of presence, showing the sense of being there with other persons.

Apart from the subjective evaluations of user experience, some objective measures of user engagement in online communities also imply significant user experience qualities in virtual gatherings. For example, Poirier-Poulin (2020) suggested that the engagement of viewers in live-streaming contributes to fostering an active atmosphere that impacts the experiences of both streamers and viewers. This dynamic is reflected in the number and content of user-generated posts (Haimson & Tang, 2017), such as the number DanMu. Analyzing log data of users' crowdsourcing control allows objective measures of their performance in tasks and would yield valuable insights into their user experience in virtual gatherings.

3. System design and implementation

3.1. Constructing a virtual world in Minecraft

Minecraft is a sandbox game that allows users to build 3D virtual worlds from scratch using blocks. We created a digital replica of a local campus in Minecraft, covering an area of 269,121 square meters. The relative size and ratio are consistent with the actual campus layout on the map. Figure 1 shows some comparative views of the campus in the real world and the Minecraft virtual world. Within Minecraft virtual worlds, players could move a character using keyboard controls: press the WASD keys to move forward, step left, move backward, and step right, and press the space key to jump. They could move the mouse to pan the viewing perspective, left click to attack, and right click to place a block and interact with objects.

3.2. DanMu control in Bilibili

Bilibili[1] is one of the most popular video sharing and streaming websites in China. It is known for its DanMu commenting system, which displays users' comments by overlaying them onto the video, moving from the right to the left (Zhou et al., 2019). To realize multiuser control based on DanMu, we streamed our local Minecraft environment to Bilibili and captured users' real-time DanMu from Bilibili.

The character movement control in Bilibili is based on keyboard commands. Participants could enter letters `W`, `A`, `S`, or `D` in the DanMu chatbox (see Figure 2, right) to move the character forward, left, backward, or right. The change of viewing perspective is not enabled with the DanMu control.

We developed three collective control methods: First Come First Served (FCFS), Vote (VOTE), and Super

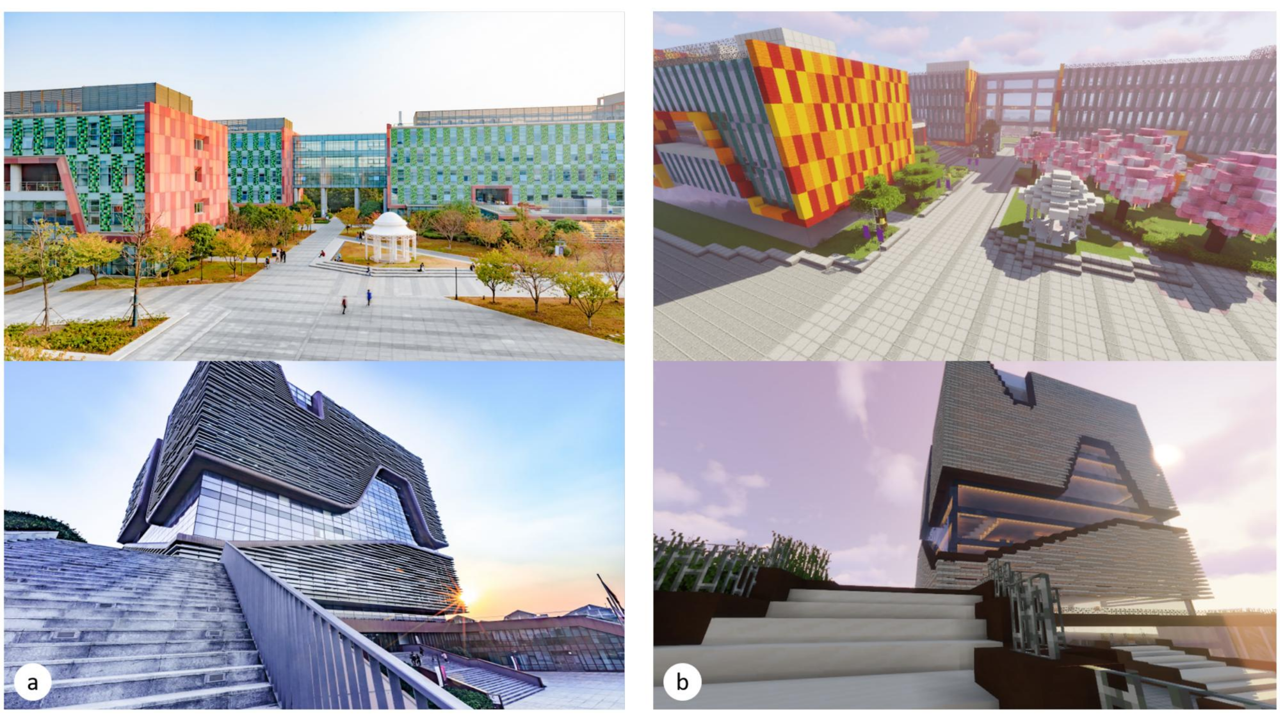


Figure 1. Comparative views of the campus in (a) the real world and (b) the Minecraft virtual world.

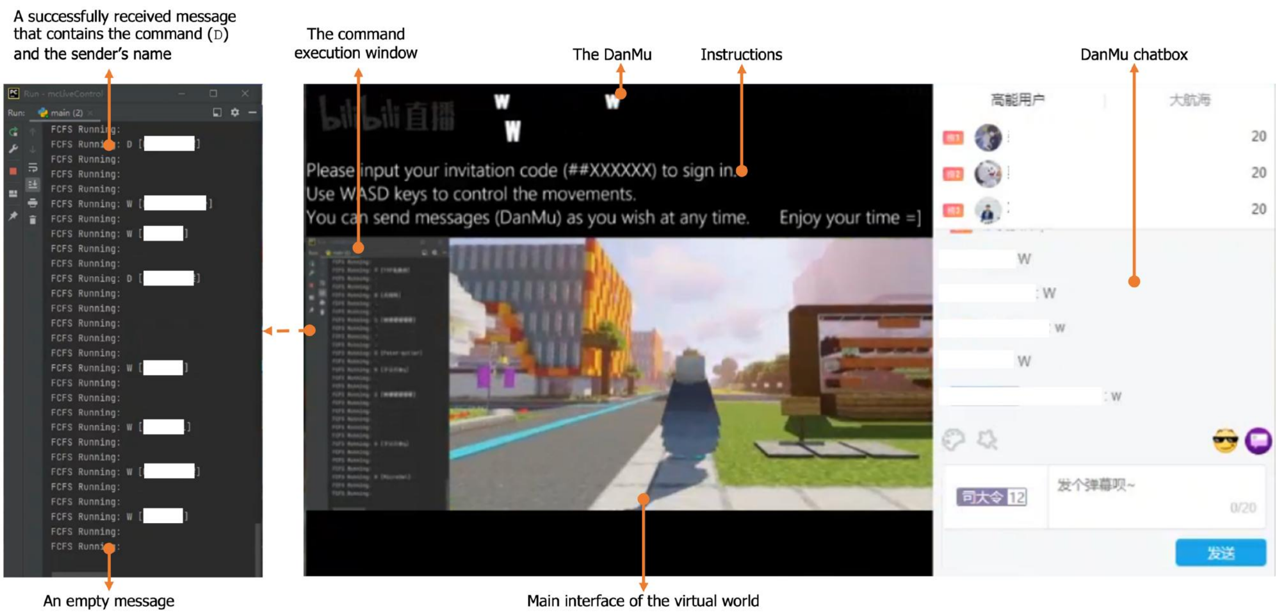


Figure 2. A Screenshot of the minecraft virtual world on bilibili, showing the command execution window (left), the virtual world (Middle), and the DanMu chatbox (right).

Table 1. Example commands received per second and the executed command based on the three control methods: First Come First Served (FCFS), vote (VOTE), and Super command (SCMD).

Example	FCFS	VOTE	SCMD
`[W, S, S]`	`W`: Forward, 1 step	`S`: Backward, 1 step	`S`: Backward, 1 step
`[W, A, S, D]`	`W`: Forward, 1 step	`W`: Forward, 1 step	`W`: Forward, 1 step
`[A, D, D]`	`A`: Left, 1 step	`D`: Right, 1 step	`D`: Right, 1 step
`WGO5`	No movement	No movement	`WGO5`: Forward, 5 steps

Command (SCMD). Table 1 shows some example executions of the three control methods.

1. **FCFS** control method executes the first command that the server receives per second.
2. **VOTE** control method executes the command that is sent the most per second. In the case of commands with the same counts, the first one is executed.
3. **SCMD** control method is based on the VOTE method. It executes the command that is sent the most per second. Aside from this, the user who sends the most DanMu (movement + communication) every 30 seconds will become the commander and is enabled to execute a super command. The command syntax is a direction letter + `GO` + a number between 1 and 9. For example, `WGO5` will trigger a continuous forward movement of 5 steps.

For each control method, participants can see the executed command in the command execution window (see Figure 2, left).

3.3. Apparatus and technical specification

Our system and programs run on a PC with an AMD Ryzen 7 5800X processor, 32 GB RAM, an NVIDIA GeForce 3060 graphics card, and a network speed of 100 Mbps. During the experiment, we captured the character's movement data on the Minecraft local server per 0.5 seconds. OBS Studio was used to live-stream two windows: the Minecraft game view and the command execution window. The command window prints a line per second, indicating the current command being executed. Given various network conditions, it is common to have ~3-4 seconds of delay in the streaming view.

To achieve the DanMu control of the character in Minecraft, we first crawled users' DanMu data using the Python `request` command. Bilibili server imposes limitations on the frequency and number of access attempts to prevent malicious traffic attacks without publicly disclosing the specific restrictions. Based on multiple tests and experimentation, we determined an access frequency of once per second. The crawled data contains all DanMu messages along with senders' IDs and timestamps, which were stored in a `JSON` file. We used regular expressions to extract command data information based on the command syntax. The data were processed so that the commands were not affected by typos. For example, `W`, `w`, and `wwwww` were all translated as a single forward movement command. To ensure the timeliness and reliability of the DanMu control, we processed the DanMu every second based on the timestamp data and implemented a queue of commands. The execution of the commands in Minecraft is accomplished through `PyAutoGUI`, a Python library that simulates keyboard and mouse inputs. In this way, when a user sends a DanMu in Bilibili, we could process it to determine if it is a command and execute a valid command to control character movements in Minecraft. On the user's side, they could see their DanMu being sent to the chatbox and floating on the main interface of the virtual world (see Figure 2). They could also see the character moving and

Table 2. Explicit user experience measured in this study, rated on a seven-point Likert scale.

	Questions	Ratings
PR1	I found the movement control in the virtual world …	1 Confusing-Structured 7
PR2		1 Impractical-Practical 7
PR3		1 Complicated-Simple 7
PR4		1 Unpredictable-Predictable 7
HE1	I found the movement control in the virtual world …	1 Dull-Captivating 7
HE2		1 Tacky-Stylish 7
HE3		1 Cheap-Premium 7
HE4		1 Unimaginative-Creative 7
UB1	This system's capabilities meet my requirements.	1 Strongly disagree – Strongly agree 7
UB2	This system is easy to use.	1 Strongly disagree – Strongly agree 7
NT1	How mentally demanding was the task?	1 Low – High 7
NT2	How physically demanding was the task?	1 Low – High 7
NT3	How hurried or rushed was the pace of the task?	1 Low – High 7
NT4	How successful were you in accomplishing what you were asked to do?	1 Good – Poor 7
NT5	How hard did you have to work to accomplish your level of performance?	1 Low – High 7
NT6	How insecure, discouraged, irritated, stressed, and annoyed were you?	1 Low – High 7
BO	It felt like I was controlling the movement of my body.	1 Strongly disagree – Strongly agree 7
AG1	It felt like I was in control of the movements during the task.	1 Strongly disagree – Strongly agree 7
AG2	What is the degree of control you felt?	1 Low – High 7
AG3	How much it felt like you are controlling the movement?	1 Not at all – Very much 7
TIME	How long do you think it took to complete the task (in minutes)?	
SR1	Please select the number that best describes your evaluation of the virtual world experience:	1 Remote – Immediate 7
SR2		1 Unemotional – Emotional 7
SR3		1 Unresponsive – Responsive 7
SR4		1 Dead – Lively 7
SR5		1 Impersonal – Personal 7
SR6		1 Insensitive – Sensitive 7
SR7		1 Unsociable – Sociable 7
PRES	In the virtual world, I had a sense of 'being there'.	1 Strongly disagree – Strongly agree 7

the command being executed in the command execution window.

4. Study 1: Crowdsourcing control methods using DanMu

4.1. Study design

In the first study, we ask **RQ1:** *How do multiplayer single-character control methods influence task efficiency and user experience in virtual gatherings?* We conducted a between-subjects study with three experimental conditions: FCFS, VOTE, and SCMD. The independent variable is the multiplayer single-character control method, and the dependent variables are task efficiency and user experience in virtual gatherings. We also measured users' use of DanMu to obtain an in-depth understanding of the control methods.

4.2. Measures

4.2.1. Task efficiency and use of DanMu

We collected user activity data during the experimental sessions, including the time spent completing the movement task, the real-time locations of the character within the virtual world, and the DanMu sent by participants. To understand the use of DanMu, we asked participants to evaluate their frequency of use after each session on a 5-point Likert scale: (1) self-reported DanMu commands: *how frequently were you sending DanMu to control the character movement?* and (2) self-reported DanMu communication posts: *how frequently were you sending DanMu to communicate with others?*

4.2.2. Explicit user experience

We refer to Bergström et al.'s study (2022) and adopted explicit measures of user experience. Table 2 includes the details. Specifically, the measures include users' pragmatic experience and hedonic experience (Hassenzahl, 2004), usability (Lewis, 2018), workload (Hart & Staveland, 1988), body ownership and agency (Bergström et al., 2022; Longo & Haggard, 2009), and time perception. A seven-point Likert scale was adopted for the questionnaire measures, except that for time perception, users were asked to enter an amount of the estimated time in minutes.

In addition, users' perceived presence was measured to reflect user experience in virtual worlds. Specifically, we included seven items that measured perceived social richness (Lombard et al., 2011). We also measured the overall presence using the classic item - *In the virtual world, I had a sense of "being there"* (Slater et al., 1994).

4.3. Experimental task and procedure

We sent out invitation emails, each with a unique code, to participants who voluntarily signed up for the study. The code consists of two hashtags and six numbers (eg, ##123456). Figure 3 illustrates the detailed experimental procedure.

At the beginning of the experiment, we explained the purpose of the study and demonstrated the use of DanMu to control character movements in Minecraft. Participants then read the information sheet and provided their consent. Apart from the collection of demographic information, participants were asked to test their network speed and fill in the results in the pre-experiment questionnaire. Two online platforms were used for the study. We used a video

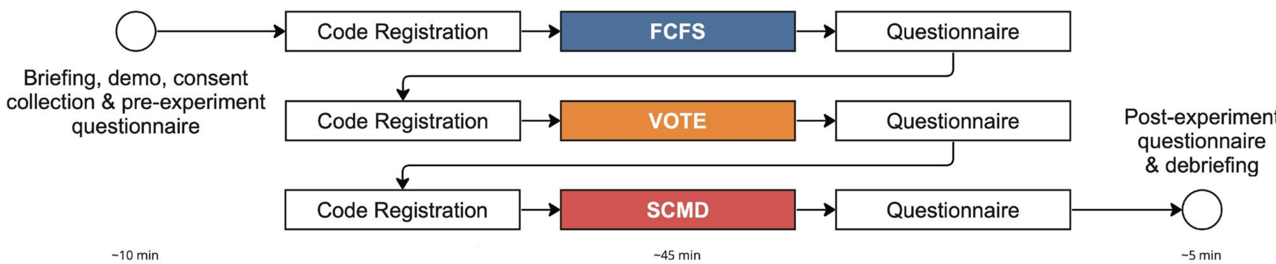


Figure 3. The experimental procedure of Study 1.

conferencing application to host and moderate the study. Participants could unmute themselves and ask questions related to the experiment at any time. In the meantime, participants were asked to join a live-streaming room on Bilibili for the three experimental sessions: FCFS, VOTE, and SCMD. They could register at any session during the live-streaming.

At the start of an experimental session, participants could see "BEGIN" in the DanMu chatbox, and they could enter their unique code to join the session (see Figure 2). Once the session was started on the server side, code registration was disabled. The task of each session was the same, and consistent verbal instructions were given to participants: *Please move the Minecraft character from the teaching building of the north campus to the cafeteria of the south campus. The blue lines indicate the recommended path, but you may also choose other paths. You can communicate the path selection using DanMu at any time. Please complete the task as soon as possible.*

During the experiment, participants were asked to mute themselves in the video conferencing software and to communicate using DanMu only. We also suggested them to turn on the sound of Bilibili live-streaming so that they could hear the background music. After each session, participants filled in a post-session questionnaire that measured their experience. The questionnaire also included two open-ended questions for qualitative feedback on the pros and cons.

The study concluded with a post-experiment questionnaire, where participants who engaged in all sessions were invited to provide a ranking of the three control methods, explain the reasons, and suggest improvements. We thanked the participants for their time, debriefed the study, and answered their questions. In total, the study took ~60 min. The study is approved by the University Ethics Committee at Xi'an Jiaotong-Liverpool University.

4.4. Participants

We sent out invitations to 240 students who were enrolled in an undergraduate module. All participants are university students with a mixed background from two schools (engineering and business) and three different majors (information and computing science, digital media and technology, and information management and information systems). In total, we had 193 participants registered in the three sessions of FCFS ($N = 78$), VOTE ($N = 64$), and SCMD ($N = 51$). However, not all registered users in the experiment filled in the demographic and user experience questionnaires. The demographic questionnaire results showed a mean age of 21.37 ($SD = 1.00$), with 58 males, 19 females, and 2 prefer not to say. Participants reported that they often use a laptop ($N = 70$), mobile phone ($N = 31$), tablet ($N = 21$), and desktop ($N = 15$) to watch online videos. Except for one participant who has never used Bilibili before, most of them ($N = 78$) are Bilibili users. The majority of them ($N = 70$) have used Bilibili for more than two years. Rated on a 5-point Likert scale, participants reported that they are moderately likely to send a DanMu when watching videos ($M = 2.68, SD = 1.25$). They were somewhat familiar with 3D images ($M = 3.10, SD = 1.31$) and experienced in gameplay ($M = 3.82, SD = 1.12$). Participants were physically located in different places on the experiment day, so their network conditions varied significantly ($M = 104.51, SD = 130.19$ for download speed; $M = 41.34, SD = 32.92$ for upload speed). The median speeds for download and upload were 83.9 Mbps and 32 Mbps, respectively. Regarding the post-session questionnaires, we received 134 complete responses for the three sessions of FCFS ($N = 64$), VOTE ($N = 42$), and SCMD ($N = 28$). Among all participants, 22 engaged in all sessions and provided their comparative evaluations.

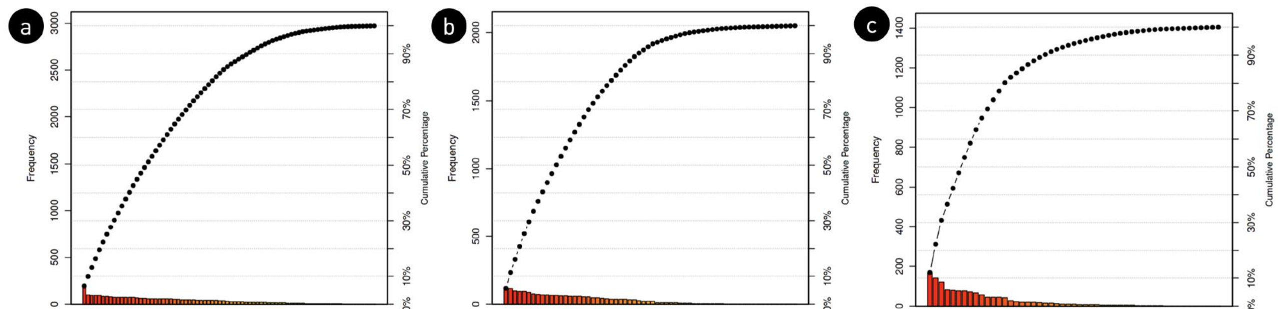
4.5. Results

During the experimental sessions, we collected user activity data, including the time spent in three sessions, the real-time locations of the character within the virtual world, and the DanMu sent by participants. For the questionnaire data, we conducted one-way ANOVA for comparisons that meet the parametric test assumptions and used Kruskal–Wallis H tests or Friedman tests otherwise for between-groups and within-groups comparisons, respectively. Bonferroni-adjusted significance values are reported for pairwise comparisons. This section includes the analysis results of the log data as well as participants' subjective evaluations.

4.5.1. Task efficiency and use of DanMu

4.5.1.1. Task completion time and real-time movements. Participants' task completion time was the shortest for SCMD (11″04′), followed by VOTE (14″00′) and FCFS (16″36′). Figure 4 shows heatmaps of the three sessions. The chosen path converged almost perfectly with the suggested path when using the SCMD, but showed deviations when using the FCFS and VOTE.

Figure 4. Heatmaps showing the character's real-time movements using three control methods: (a) First Come First Served (FCFS), (b) vote (VOTE), and (c) Super command (SCMD).



(a)

	Max	Min	Mean (SD)	Sum
Total N	195	1	38.08 (34.01)	2970
N / minute	11.75	0.06	2.29 (2.05)	178.92

(b)

	Max	Min	Mean (SD)	Sum
Total N	117	1	32.05 (32.96)	2051
N / minute	8.36	0.07	2.29 (2.35)	146.50

(c)

	Max	Min	Mean (SD)	Sum
Total N	169	1	27.53 (38.74)	1404
N / minute	15.27	0.09	2.49 (3.50)	126.87

Figure 5. Pareto charts and descriptive tables showing the frequency distributions of DanMu commands in (a) First Come First Served (FCFS), (b) vote (VOTE), and (c) Super command (SCMD).

4.5.1.2. Time perception. A one-way ANOVA showed a statistically significant difference in time perception among the three control methods, $F(2, 13) = 3.61, p = 0.03$. Post-hoc pairwise comparisons showed that participants perceived significantly shorter time using SCMD ($M = 12.39, SD = 5.39$) than FCFS ($M = 16.25, SD = 6.24$), $p = 0.025$.

4.5.1.3. Self-reported DanMu commands. A Kruskal–Wallis H test showed no statistically significant difference in participants' self-reported DanMu commands among three control methods, $\chi^2(2) = 0.089, p = 0.957$, with a mean rank of 68.48 for FCFS, 66.31 for VOTE, and 67.05 for SCMD.

4.5.1.4. Self-reported DanMu communication posts. A Kruskal–Wallis H test showed no statistically significant difference in participants' self-reported DanMu communication posts among three control methods, $\chi^2(2) = 0.44, p = 0.802$, with a mean rank of 65.73 for FCFS, 70.45 for VOTE, and 67.11 for SCMD.

4.5.1.5. Logged DanMu Counts. A one-way ANOVA showed no statistically significant difference in logged DanMu counts among the three control methods, $F(2, 190) = 1.46$, $p = 0.23$. Figure 5 shows the ordered frequency of DanMu sent by participants when using the three control methods. The cumulative percentage lines showed that a small number of users were particularly active when using SCMD: six users (12%) contributed to nearly half of the total DanMu. Comparatively, 12 users (19%) and 17 users (22%) contributed to half of the DanMu in VOTE and FCFS, respectively. The normalized frequency showed that on average, participants sent the greatest number of DanMu per minute when using SCMD ($M = 2.49, SD = 3.50$).

4.5.2. Explicit user experience

Figure 6 shows a summary of the user experience and perceived presence reported by participants in post-session questionnaires.

4.5.2.1. Pragmatic experience. A Kruskal–Wallis H test showed a statistically significant difference in pragmatic experience among three control methods, $\chi^2(2) = 7.47, p = 0.024$, with a mean rank of 60.34 for FCFS, 67.35 for VOTE, and 84.11 for SCMD. Post-hoc pairwise comparisons showed that participants had significantly greater pragmatic experience using SCMD than FCFS ($p = 0.019$).

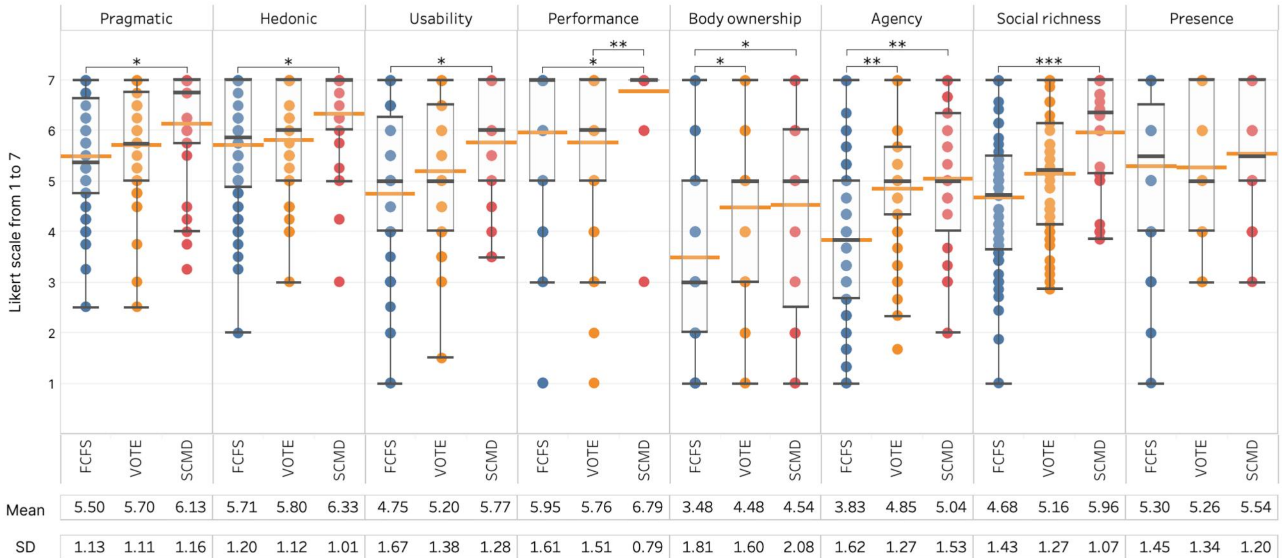


Figure 6. Box plots with means and standard deviations showing between-groups comparisons of participants' self-reported user experience and presence using three control methods ($N = 134$). FCFS: First Come First Served; VOTE: vote; SCMD: Super command.

4.5.2.2. Hedonic experience. A Kruskal–Wallis *H* test showed a statistically significant difference in hedonic experience among three control methods, $\chi^2(2) = 6.69, p = 0.035$, with a mean rank of 62.20 for FCFS, 64.70 for VOTE, and 83.80 for SCMD. Post-hoc pairwise comparisons showed that participants had significantly greater hedonic experience using SCMD than FCFS ($p = 0.035$).

4.5.2.3. Usability. A Kruskal–Wallis *H* test showed a statistically significant difference in pragmatic experience among three control methods, $\chi^2(2) = 8.10, p = 0.017$, with a mean rank of 59.35 for FCFS, 68.90 for VOTE, and 84.02 for SCMD. Post-hoc pairwise comparisons showed significantly greater usability in SCMD than in FCFS ($p = 0.014$).

4.5.2.4. Workload. A Kruskal–Wallis *H* test showed no statistically significant difference in mental, physical, temporal demand, effort, or frustration. However, there was a statistically significant difference in performance, $\chi^2(2) = 13.94, p = 0.001$, with a mean rank of 64.56 for FCFS, 55.94 for VOTE, and 86.77 for SCMD. Post-hoc pairwise comparisons showed that participants had a significantly greater sense of performance using SCMD than using FCFS ($p = 0.013$) and VOTE ($p = 0.001$).

4.5.2.5. Body ownership. A Kruskal–Wallis *H* test showed a statistically significant difference in body ownership among three control methods, $\chi^2(2) = 10.18, p = 0.006$, with a mean rank of 56.45 for FCFS, 76.99 for VOTE, and 78.52 for SCMD. Post-hoc pairwise comparisons showed that participants had significantly greater body ownership using VOTE ($p = 0.021$) and SCMD ($p = 0.034$) than using FCFS.

4.5.2.6. Agency. A Kruskal–Wallis *H* test showed a statistically significant difference in agency among three control methods, $\chi^2(2) = 15.59, p < 0.001$, with a mean rank of 53.79 for FCFS, 78.32 for VOTE, and 82.61 for SCMD. Post-hoc pairwise comparisons showed that participants had significantly greater agency using VOTE ($p = 0.004$) and SCMD ($p = 0.003$) than using FCFS.

4.5.2.7. Social richness. A Kruskal–Wallis *H* test showed a statistically significant difference in social richness among three control methods, $\chi^2(2) = 15.98, p < 0.001$, with a mean rank of 56.27 for FCFS, 68.79 for VOTE, and 91.25 for SCMD. Post-hoc pairwise comparisons showed that participants had significantly greater social richness using SCMD than FCFS ($p < 0.001$).

4.5.2.8. Presence. A Kruskal–Wallis *H* test showed no statistically significant difference in presence among three control methods, $\chi^2(2) = 0.65, p = 0.724$, with a mean rank of 67.28 for FCFS, 64.74 for VOTE, and 72.14 for SCMD.

4.5.3. Within-groups comparisons

We further looked into the experiment data of the 22 participants who joined all three sessions and conducted a within-groups comparison of the measures (see Figure 7). Friedman's tests showed significant differences in usability ($\chi^2(22) = 7.40, p = 0.025$), time perception ($\chi^2(22) = 15.16, p = 0.001$), and social richness ($\chi^2(22) = 7.10, p = 0.029$) among the three control methods. Post-hoc pairwise comparisons showed significantly greater usability ($p = 0.016$), less perceived time ($p = 0.003$), and greater social richness ($p = 0.024$) using SCMD than FCFS.

4.5.4. Subjective ranking

A Friedman's test showed a significant difference in participants' subjective ranking, $\chi^2(2) = 7.30, p = 0.026$, with a mean rank of 2.40 for FCFS, 2.05 for VOTE, and 1.55 for SCMD. As participants rated them on the 1st, 2nd, and 3rd,

	Pragmatic FCFS	Pragmatic VOTE	Pragmatic SCMD	Hedonic FCFS	Hedonic VOTE	Hedonic SCMD	Usability FCFS	Usability VOTE	Usability SCMD	Performance FCFS	Performance VOTE	Performance SCMD	Body ownership FCFS	Body ownership VOTE	Body ownership SCMD	Agency FCFS	Agency VOTE	Agency SCMD	Social richness FCFS	Social richness VOTE	Social richness SCMD	Presence FCFS	Presence VOTE	Presence SCMD
Mean	5.89	5.75	6.09	6.03	5.99	6.22	5.00	5.25	5.73	6.27	6.32	6.73	3.73	4.59	4.45	4.21	4.82	4.86	5.29	5.49	5.93	5.86	5.45	5.41
SD	0.99	1.30	1.15	1.34	1.19	1.10	1.68	1.47	1.31	1.58	1.29	0.88	2.10	1.65	2.20	1.91	1.36	1.63	1.51	1.23	1.15	1.28	1.37	1.26

Figure 7. Box plots with means and standard deviations showing within-groups comparisons of participants' self-reported user experience and presence using three control methods ($N = 22$). FCFS: First Come First Served; VOTE: vote; SCMD: Super command.

a lower mean rank indicates a higher ranking here. Post-hoc pairwise comparisons showed that participants ranked significantly higher on SCMD than FCFS ($p = 0.022$).

4.5.5. Observations and user feedback

Our monitoring of user activity data showed that some participants were extremely active in sending DanMu commands, while some barely sent anything and chose to be passive viewers of the live-streaming. The frequency of DanMu ranged from 1 to 195 (see Figure 5). Participants endorsed SCMD, acknowledging the stimulating effect of being a commander on their DanMu engagement and the effectiveness of the representative control in their collective activities. Another notable comment from participants was that they found FCFS less favorable mostly because of the large group size and suggested a smaller size for its use. We took this into consideration when we designed Study 2.

5. Study 2: User activeness and group size

5.1. Study design

In the second study, we explored **RQ2:** *How do user activeness and group size influence task efficiency and user experience in virtual gatherings?* Because there were few differences in user experience found between VOTE and the other two control methods in Study 1, we excluded it from this study to simplify the study design and save participants' time. We designed this study with two independent variables: user activeness (active and inactive) and group size (large and small). We consider our Study 1 as a large group study because *Twitch* has an average of 27 viewers per channel (TwitchTracker, 2024), and the sample sizes of our groups in Study 1 were 2-3 times larger. To determine how small the group size should be, we found that previous work has explored crowdsourcing methods with a group of nine users (Lesniak & Maistro, 2022). Ranking the number of DanMu commands participants sent in Study 1, we found that the top ten active users have contributed to 30%–70% of the DanMu numbers, and the bottom ten users were largely inactive in sending DanMu. Thus, we invited the top and bottom users on the ranked list to form active and inactive groups of ten, respectively (see Figure 5). The measures in this study remained consistent with those used in Study 1 (see Section 4.2), with the exception that we excluded the six workload items because, aside from performance, we did not observe significant differences in the measures.

5.2. Experimental task and procedure

The experimental procedure is similar to the first study (see Section 4.3). We started with a briefing session and demonstrated the techniques. Participants' consent and their demographic information were collected. After this, we conducted two experimental sessions (FCFS and SCMD). Within each session, participants were given the same task instructions as those provided in Study 1. They were asked to collectively move the character in the live-streaming view from the teaching building of the north campus to the cafeteria of the south campus. The order of the two techniques was counterbalanced by following the Latin Square Design. The study concluded with a post-experiment questionnaire and a debriefing session with open discussions. The experiment in Study 2 took ~40 min in total.

5.3. Participants

Twenty participants rejoined the study (15 males, 5 females; aged $M = 21.6, SD = 1.35$). They were familiar with 3D images ($M = 3.95, SD = 1.32$) and experienced in gameplay ($M = 4.40, SD = 0.88$). One participant from the active

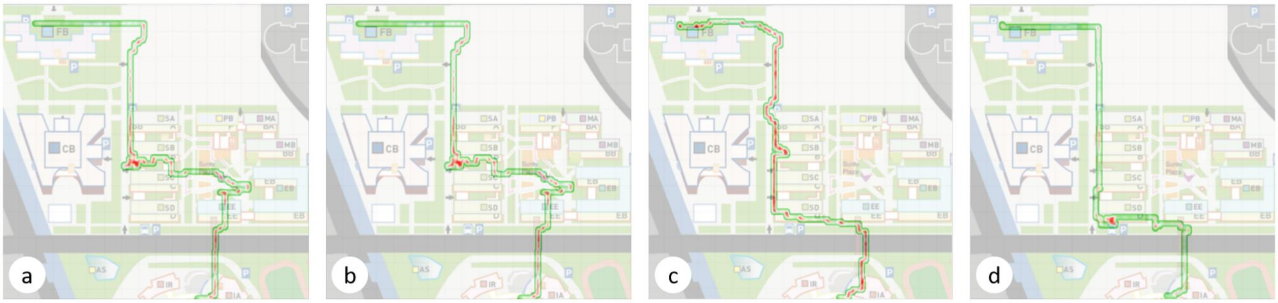


Figure 8. Heatmaps showing the character movements in (a) A-FCFS, (b) A-SCMD, (c) I-FCFS, and (d) I-SCMD. A: active; I: inactive; FCFS: First Come First Served; SCMD: Super command.

Table 3. Summary table of participants' self-reported DanMu frequency and logged DanMu counts.

		Active		Inactive	
	Group	FCFS M (SD)	SCMD M (SD)	FCFS M (SD)	SCMD M (SD)
Self-reported Command DanMu	Small	4.9 (0.3)	5 (0)	3.5 (1.3)	3.8 (1)
	Large	4.5 (1.4)	4.7 (0.8)	3.7 (1.2)	4 (1.2)
Self-reported Communication DanMu	Small	1.9 (1.3)	1.5 (1.1)	1.8 (1.7)	1.9 (1.7)
	Large	2.75 (1.4)	2.7 (2)	2.5 (1.4)	3 (2)
Logged DanMu Count	Small	234.8 (85.9)	123.5 (41.9)	39.1 (30.5)	36.9 (18.3)
	Large	81.1 (48.5)	63.6 (47.0)	44.7 (24.2)	41.8 (53.1)

group had used Bilibili for 1-2 years. The rest 19 participants had used it for over two years. Participants reported that they are moderately likely to send a DanMu when watching videos for both the active group ($M = 2.80, SD = 1.40$) and the inactive group ($M = 2.80, SD = 1.03$). Participants were physically located in different places on the experiment day, so their network conditions varied significantly ($M = 102.07, SD = 86.32$ for download speed; $M = 35.50, SD = 23.29$ for upload speed). The median speeds for download and upload were 87.29 Mbps and 31.2 Mbps, respectively.

5.4. Results

We conducted *t* tests to compare the differences between active and inactive groups, two control methods, and large and small groups. Mann Whitney *U* tests and Wilcoxon Sign tests were used, and the mean rank (MR) values were reported for comparisons that failed to meet the parametric test assumptions.

5.4.1. Task completion time and real-time movements

For small groups with active users, participants' task completion time was shorter using SCMD ($10''52'$) than FCFS ($15''36'$). Similar results were found for small groups with inactive users (SCMD: $9''25'$; FCFS: $10''57'$). Participants' self-reported time perceptions were roughly consistent with the logged results. The active group reported a mean time (in minutes) of 10.9 ($SD = 4.12$) for SCMD and 17.4 ($SD = 4.22$) for FCFS; the inactive group reported a mean time of 9.8 ($SD = 3.85$) for SCMD and 13.9 ($SD = 5.17$) for FCFS. Figure 8 shows heatmaps of the character movements in the four sessions. It is worth noting that the active group deliberately took a longer path (see Figure 8a,b). During the experiment, they communicated and reached an agreement to explore different views of the campus. On the other hand, the inactive group successfully followed the suggested path and completed the movement tasks.

5.4.2. Active and inactive groups

5.4.2.1. Explicit user experience. Mann Whitney *U* tests showed that participants in the active group reported significantly greater hedonic experience ($Z = -2.25, p = 0.033, MR(A) = 24.45, MR(I) = 16.55$), and social richness ($Z = -2.44, p = 0.014, MR(A) = 25.00, MR(I) = 16.00$). Differences in other aspects of user experience were statistically insignificant (see Figure 9).

5.4.2.2. Reported and Logged DanMu Counts. Mann Whitney *U* tests showed that participants in the active group reported significantly greater use of DanMu commands ($MR = 27.22$) than the inactive group ($MR = 13.78$), $Z = -4.20, p < 0.001$. A between-groups *t* test showed that users in the active group sent a significantly larger number of DanMu ($M = 179.15, SD = 58.89$) than users in the inactive group ($M = 38, SD = 23.4$), $t(12) = 7.04, p < 0.001$. Table 3 shows the details about participants' self-reported DanMu frequency and logged DanMu counts.

5.4.3. FCFS and SCMD

5.4.3.1. Explicit user experience. Similar to the results in Study 1, participants reported significantly greater usability($Z = 2.21, p = 0.03, MR(FCFS) = 16.52, MR(SCMD)$- $Z = 2.21, p = 0.03, MR(FCFS) = 16.52, MR(SCMD) = 24.48$) and less perceived time ($Z = -3.30$, $p = 0.001, MR(FCFS) = 26.48$, $MR(SCMD) = 14.52$) using SCMD than FCFS. However, the difference in social richness became

	Pragmatic Active	Pragmatic Inactive	Hedonic Active	Hedonic Inactive	Usability Active	Usability Inactive	Body ownership Active	Body ownership Inactive	Agency Active	Agency Inactive	Social richness Active	Social richness Inactive	Presence Active	Presence Inactive
Mean	6.36	6.09	6.45	5.50	5.73	5.70	5.00	4.10	5.22	5.10	6.04	4.97	5.95	5.30
SD	0.69	1.12	0.61	1.33	1.27	1.37	1.95	1.97	1.10	1.27	0.63	1.31	0.76	1.49

Figure 9. Box plots with means and standard deviations showing between-groups comparisons of participants' self-reported user experience and presence in active and inactive groups ($N = 40$).

insignificant in small groups, $Z = 0.60, p = 0.57$ (see Figure 10). All participants in the active group preferred SCMD ($N = 10$) to FCFS. On the other hand, three participants in the inactive group preferred FCFS, and one participant reported no difference.

5.4.3.2. Reported and Logged DanMu Counts. Participants' self-reported use of DanMu commands showed no significant differences for FCFS and SCMD. However, the analysis of the DanMu logs showed that active participants sent a significantly larger number of DanMu using FCFS ($M = 234.8, SD = 85.9$) than using SCMD ($M = 123.5$, $SD = 41.9$), $t(9) = 5.30, p < 0.001$. However, there was no significant difference for inactive users using FCFS ($M = 39.1, SD = 30.5$) and SCMD ($M = 36.9$, $SD = 18.3$), $t(9) = 0.37, p = 0.72$.

5.4.4. Large group and small group

5.4.4.1. Explicit user experience. We compared data collected from this study with the large-group results in Study 1. Mann Whitney U tests showed that participants in the small group reported significantly greater pragmatic experience ($Z = -2.68, p = 0.007, MR(S) = 105.91$, $MR(L) = 82.00$), usability ($Z = -2.22, p = 0.027, MR(S) = 102.81$, $MR(L) = 82.93$), and agency ($Z = -2.87$, $p = 0.004$, $MR(S) = 107.50, MR(L) = 81.53$). Differences in other aspects of user experience were statistically insignificant (see Figure 11). Within-group comparisons of Study 1 and 2 showed no statistically significant difference in user experience. Most participants ($N = 19$) preferred the small group experience, whereas only one participant preferred the large group experience.

5.4.4.2. Reported and Logged DanMu Counts. Mann Whitney U tests showed that participants in small groups reported significantly greater use of DanMu commands ($MR = 111.30$) than large groups ($MR = 80.40$), $Z = -3.55, p < 0.001$. Additionally, a Wilcoxon signed-rank test showed a significant decrease in users' self-reported communication DanMu in small groups compared to large groups, $Z = -3.064, p = 0.02$. The analysis of the DanMu logs showed that active participants in a small group ($M = 179.15, SD = 58.98$) sent a significantly larger number of DanMu than in a large group ($M = 72.35$, $SD = 32.44$), $t(9) = -5.58, p < 0.001$. However, there was no significant difference for inactive users in large ($M = 43.25, SD = 26.18$) and small groups ($M = 38$, $SD = 23.40$), $t(9) = -0.45, p = 0.66$.

6. Discussion

6.1. Effects of control methods on task efficiency and user experience

Our first study answered **RQ1:** *How do multiplayer single-character control methods influence task efficiency and user experience in virtual gatherings?* Participants' task completion time showed that the democracy modes (SCMD and VOTE) outperformed the anarchy mode (FCFS) in task efficiency. This supports our **H1a**. In addition, we found that having a representative in SCMD further contributes to the task efficiency of VOTE. In terms of user experience, we found significant effects of control methods on pragmatic and hedonic user experience, usability, perceived performance, body ownership, agency, and social richness. In particular, participants reported significantly greater user experiences using SCMD than FCFS. Our **H1b** was not supported. The main reason perhaps lies in the difference between experimental tasks and gameplay. Our experimental tasks lasted for a short period of time and demanded task efficiency, whereas in gameplay, efficiency is less emphasized, and it would normally last a longer period of time. While the unpredictability in the anarchy mode has a negative effect

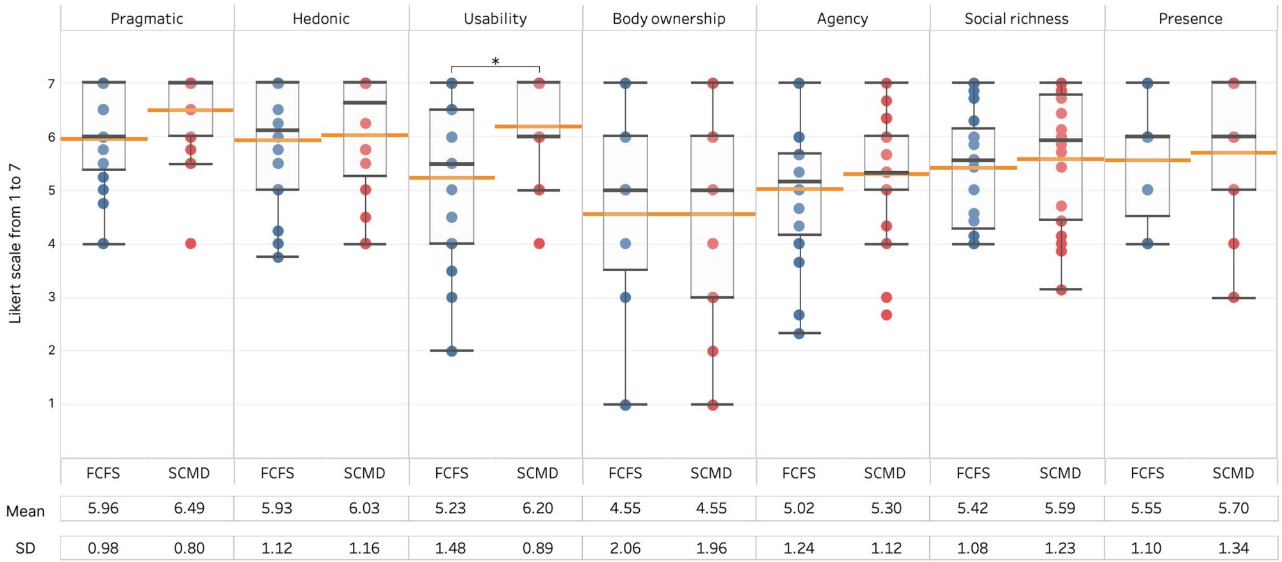


Figure 10. Box plots with means and standard deviations showing within-groups comparisons of participants' self-reported user experience and presence using two control methods ($N = 20$). FCFS: First Come First Served; SCMD: Super command.

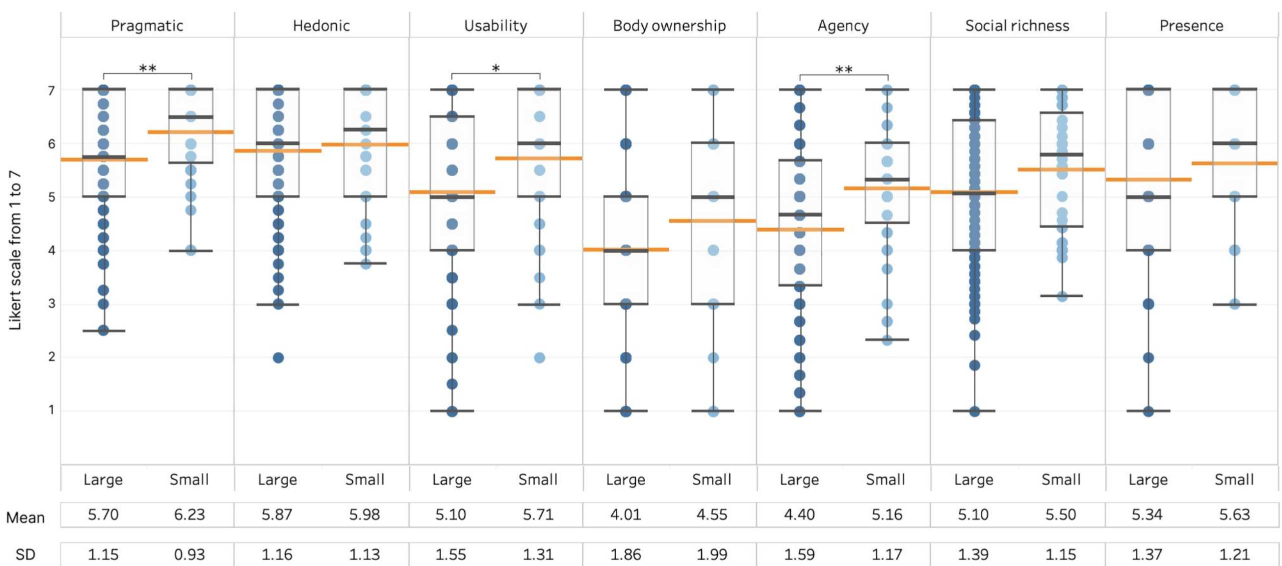


Figure 11. Box plots with means and standard deviations showing between-groups comparisons of participants' self-reported user experience and presence in large and small groups ($N = 174$).

on task efficiency and user experience in experimental tasks, it may, on the other hand, contribute to experience in gameplay.

Participants who preferred SCMD commented that "*a correct command from a leader could significantly increase the movement efficiency*", and they liked SCMD as it is "*quicker*", "*more efficient*", and "*offers a greater sense of participation*". In the meantime, SCMD faces a limitation in a large group that users might feel disappointed when it becomes challenging to become a commander. Participants also liked their names being shown to everyone, which gave them a sense of recognition and achievement. For example, "*I felt very engaged seeing my name on the super command*". Acknowledging users' identity is important for crowd behavior. This finding is in line with the statement of Seering et al. (2017) that the recognition in the community will largely affect the user experience in participating in gathering activities.

However, some participants preferred VOTE and identified some limitations of SCMD. For example, "*VOTE is more effective and fair, and I felt more involved. SCMD makes us fight for the commander, not to control the character*". Some participants also mentioned that "*some people just don't make good use of the super command on purpose, so I prefer VOTE*", and "*it is relatively easy to carry out the correct operation using VOTE*". The efficiency of VOTE in performing the most appropriate action could provide a sense of control in virtual gathering, which reflects the need of participants to generate meaningful results (Sakamoto et al., 2017).

FCFS was ranked the lowest. Participants mentioned that "*it is easy to go wrong because of an incorrect command in*

FCFS" and "*FCFS is more vulnerable to distraction messages*". Nevertheless, there were some participants who preferred FCFS and reported that it "*always provides a sense of engagement. I felt that I was just observing during SCMD if I couldn't get it*". On the one hand, this finding again confirms the importance of providing an effective mechanism to guide the gathering activities. On the other hand, a certain level of randomness, which could lead to a special experience of improvisational events (Lytle et al., 2020), may also provide enjoyment for participants.

Overall, our results showed that for a large group, users perceived significantly different user experiences when using the three control methods. Notably, democracy modes (SCMD and VOTE) outperformed the anarchy mode (FCFS) in terms of body ownership and agency. Among the three methods, SCMD was found to show the greatest user experience in all aspects and supported the greatest sense of social richness. The results from participants' subjective rankings and feedback indicated that it is necessary to have representative controls for MSCC during virtual gathering activities.

6.2. Effects of user activeness and group size on task efficiency and user experience

The second study investigated **RQ2:** *How do user activeness and group size influence task efficiency and user experience in virtual gatherings?* Within a small group, we found that participants required less time to complete the task using SCMD than FCFS, which is consistent with the findings in Study 1. However, active users spent a longer time than inactive users, which was an unexpected finding and failed to support the task efficiency aspect of **H2a**. The analysis of movement paths showed that active users chose a longer path while inactive users followed the suggested path. This led to the difference in task completion time. Overall, participants completed the movement task more efficiently in a small group than in a large group, supporting our **H2b**. In terms of user experience, the results showed significant effects of user activeness on hedonic experience and social richness and a significant effect of group size on pragmatic experience, usability, and agency. We discuss the detailed findings in the following sections.

The results showed that active users perceived significantly greater hedonic user experience and social richness than inactive users. The user experience aspect of **H2a** is supported. This finding is also reflected in participants' comments. Active users reported that "*it's so fun*", "*awesome*", "*I was really engaged*", "*I got the super command hahahahaha*", and "*we're so united*". Two interesting findings were observed within active and inactive groups. One was that users in the inactive group used less time in completing the task than users in the active group. Comparing (a–b) with (c–d) in Figure 8, it is clear that the active group users chose to try a different but longer path. They reported in the debriefing that they were already familiar with the shortest path so they would like to explore different campus views. Thus, it may not be a fair comparison of the task completion time when the two groups chose different paths. Nevertheless, this difference in the decision-making process shows a clear separation between the two groups, which echoes the finding in Poirier-Poulin (2020) that higher activeness of the group may stimulate the creativity of users.

The other interesting observation was about the preferred control methods. Brandis and Bozkurt (2021) found in their study that active users may prefer more democratic methods, whereas inactive users would prefer anarchy modes. However, our study showed some contradictory findings. All active users preferred SCMD over FCFS, but three participants in the inactive group preferred FCFS. They reported that "*it (FCFS) is more engaging, as the character can move under my command as long as I'm the first*", and "*I like it because I had a greater sense of participation using FCFS*". This seems consistent with participant feedback in Study 1, where participants who found it difficult to get a super command preferred not to have it.

The results also showed that users perceived greater pragmatic experience, usability, and sense of agency in small groups compared to a large group, supporting **H2b**. Participants reported that "*I like the small group because it is more efficient with fewer people and there was no one making trouble*". Within a small group, participants found that "*it's easier to get the super command because there are much fewer people*". Similar findings were also observed in existing studies (eg, Kim, 2013; Shaw, 2013), in which the smaller group size is positively related to the active participation and experience of individuals. However, one participant preferred a large group, commenting that "*there were more people and better communication last time. This time, we were more focused on completing the task, and there was barely any communication*". The importance of communication in large groups was also recognized in previous works (eg, Deng et al., 2015; Oldenburg, 1999), indicating the key role of conversations in online virtual spaces.

To summarize, active users are keener on the efficiency of collective control, but some inactive users tend to prefer to observe and value the sense of communication in collective behaviors. The hedonic experience and social richness varied among the two types of users. Users generally prefer to work in a small group on tasks that demand efficiency, as the sense of agency and the pragmatic experience of collective control were perceived higher.

6.3. Command DanMu and communication DanMu

We asked users how often they sent commands and communication DanMu during the experiment, along with explanations. Participants who reported frequent command DanMu explained that they were highly engaged because they want to "*be the top one*", "*complete the tasks*", "*achieve the goal*", and "*have fun*". One participant mentioned that "*I was trying to send more correct commands to avoid the chance of incorrect command being executed*". While most participants were very task-driven and goal-oriented, several participants reported non-collective behaviors. For example, "*I want to see more about the virtual campus, so I was just* *typing random commands"*, and *"I was trying to control the avatar and also making some troubles"*. For participants who reported that they did not send any command DanMu, they explained that they *"prefer to observe"*, and they found it *"not necessary"*. The non-collective behaviors were observed less in a small group, but inactive behaviors remained the same in both large and small groups, indicated by the significantly smaller number of DanMu they sent.

Comparatively, communication DanMu was much less than command DanMu. This is mainly because participants found it unnecessary, too difficult, or not responsive. Most participants reported that *"it is mainly the control of the character, not to communicate"*, and *"we are very collective so it was not necessary"* to send communication DanMu. Some mentioned that they were *"too busy (in control) to communicate"*, and that *"there were too many commands and it's difficult to see what I say"*. One participant mentioned that *"I wanted to communicate with others, but no one replies, so I gave up"*. While participants reported moderately frequent use of communication DanMu in a large group, the frequency dropped when they were in a small group gathering. One participant reported that the need to switch keyboard languages to Chinese hindered the use of communication DanMu, indicating a potential socio-cultural influence on the user experience. One of the few participants who reported a very frequent use of communication DanMu mentioned that *"I want to provide instructions to everyone's movement, rather than making the movement myself"*. The overall communication DanMu remained low, and it seems to indicate that communicating via typing is not ideal for collective activities using command-based control methods. Other approaches, such as voice messages, may better facilitate communication.

6.4. Design guidelines, implications, and future work

Our analysis points to the following design guidelines for **crowdsourcing control**:

1. **Task Efficiency.** For tasks that necessitate repetitive execution of commands and demand efficiency, consider adopting a democracy mode with representative controls (eg, Super Command). This avoids the uncertainty and unexpected results from the anarchy mode (eg, FCFS). Participants' task completion time using three control methods gave strong implications for this.
2. **Group Size.** Use democratic control for optimal user experience in large groups. We suggest this because crowdsourcing control methods were found to have significant impacts on various aspects of user experience in a large group, among which only usability was found significantly higher using SCMD in a small group. Small groups allow for more flexible use of crowdsourcing control methods.
3. **User Activeness.** Users who are actively engaged in collective activities prefer democratic modes of control for compelling tasks ($\sim$10 minutes). However, it is likely to show contradictory results for experience spanning a longer time, as some users still appreciate the anarchy mode and find it highly engaging.

The following lessons are learned regarding **user experience in virtual gatherings** based on DanMu:

1. **Show Credits.** Give credit to users by indicating their names along with the active commands that affect the collective behaviors. Our participants' feedback confirmed its contribution to their user engagement in virtual gatherings.
2. **Performer and Spectator.** Be aware of different user types in virtual gatherings. Prioritize users' active participation to support collective experience. In the meantime, allow users to be spectators and facilitate their communication.
3. **User Engagement.** Users tend to be more active and prefer small groups ($\sim$10 users) over large groups ($>$50 users) in virtual gatherings based on DanMu. Support small-group activities if a strong sense of agency and pragmatic quality are desired.

Our results and findings support the use of cost-effective, widely accessible, and easy-to-learn approaches (ie, Bilibili and DanMu) to engage users in collective activities in an online virtual gathering. The current work focused on multiplayer single-character control, which can be generalized to many application areas in future metaverse that necessitate social interactions, focused attention, and collective experiences. The findings are also applicable to general 3D virtual environments other than Minecraft. Here we name a few example scenarios that can apply the design guidelines in practice. For example, in a group virtual tour, visitors could collectively determine a travel path. In this case, we recommend that users should be allowed to vote if it is a large group ($>$50), so that users can have a greater sense of agency in control and social richness. For an immersive virtual performance, audiences can have a participatory experience of the show, travel around the scene, and unlock new plots. In this scenario, adopting FCFS may introduce a sense of uncertainty and enrich the experience, as efficiency is likely not the primary concern. In the meantime, it is necessary to allow users to be spectators. In a virtual classroom, collective control allows students in a group to determine whether or not to move forward to the next part of the teaching and when to do so. Showing credits to students who contributed to the teaching progress in this scenario may stimulate students' motivation to learn. These example scenarios cover typical use cases of future metaverse in tourism, entertainment, and education.

The current work has some limitations. First, while our studies have reasonable sample sizes, it is not representative of the general public. Participants of our study are mainly university students, and most of them are males. Given it is a virtual gathering in a campus environment, our study findings could implicate future work on gathering activities taking place in an educational environment and designed for young adults. However, researchers and designers should be aware of the potential influence of demographic characteristics (eg, experience in gameplay and socio-cultural background) on user behaviors, as indicated in previous work (Monteiro et al., 2022). Second, there was a noticeable delay in the streaming view and the active commands. However, some participants reported that they had managed to gauge the streaming delay and adapted their commands to it, which significantly increased their experience. Third, the constructed environment in Minecraft simulates relatively simple control tasks (ie, movement only). Future work should explore more complex MSCC tasks, such as the change of views, the manipulation of virtual objects, and even social interactions with virtual avatars. Allowing users to define new command syntax for customized controls would also be an interesting extension in future work. This feature will enable adaptable DanMu controls based on user preferences and support the optimization of DanMu controls on the go. While we did not aim to improve the efficiency or fairness of crowdsourcing control methods, our work can be combined with existing works to assign weighting to players' inputs when aggregating the commands (Lesniak & Maistro, 2022). Finally, the virtual gathering in our experiment was largely task-driven and time-intensive as we were interested in understanding the efficiency of crowdsourcing control methods. Virtual gatherings in the metaverse are likely to happen on more casual occasions and around complex tasks, such as giving presentations in lectures and having brainstorming activities in creative design. The examination of the social dynamics associated with the use of DanMu as a communication tool is poised to yield further insights and warrants investigation in future research.

7. Conclusion

Web-based virtual worlds are likely to be the near-future medium of the metaverse. This work examined a novel approach that allows groups of online users to interact with a live-streamed virtual world using DanMu. Specifically, we studied three methods for multiplayer single-character control (MSCC): First Come First Served (FCFS), Vote (VOTE), and Super Command (SCMD). Our results showed that in a large group (>50 users), SCMD allowed users to perceive significantly greater user experience and social richness than FCFS, but the presence was not affected by the collective control methods. These findings provide answers to **RQ1**. In small groups (~10 users), the differences in user experience using varying control methods were insignificant, except that SCMD had higher usability. Virtual gatherings in a small group allowed greater pragmatic quality, usability, and a sense of agency than in a large group, but the hedonic quality, social richness, and presence did not significantly differ. In addition, active users perceived greater hedonic experience and social richness than inactive users. These findings provide answers to **RQ2**. We validated three DanMu controls for multiplayer single-character control and suggested design guidelines for future work. Our work showed that DanMu not only enables communication and sharing reactions in online video sharing platforms, but can also be used as a control method to facilitate collective experiences in virtual worlds. Despite the opportunities to enhance social connections through DanMu in future virtual gatherings, challenges remain in effectively moderating DanMu and addressing privacy concerns in a growing and diverse user base. It is also necessary to consider the appropriateness and needs of the virtual world's context when designing systems based on DanMu. Our work has implications for future human-computer interaction with the metaverse in the fields of gameplay, tourism, entertainment, and education.

Note

1. https://live.bilibili.com.

Acknowledgements

We thank our participants for their time and the reviewers for their comments and suggestions that helped us improve our paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work is supported by the National Natural Science Foundation of China (#62207022), the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (#22KJB520038), and Xi'an Jiaotong-Liverpool University (#RDF-20-02-47).

ORCID

Yue Li http://orcid.org/0000-0003-3728-218X
Teng Ma http://orcid.org/0000-0002-1285-8106
Ziming Li http://orcid.org/0009-0004-7529-7176
Hai-Ning Liang http://orcid.org/0000-0003-3600-8955

Data availability statement

Data associated with this research can be made available upon reasonable request to the corresponding author.

References

Aleta, A., & Moreno, Y. (2019). The dynamics of collective social behavior in a crowd controlled game. *EPJ Data Science*, *8*(1), 22. https://doi.org/10.1140/epjds/s13688-019-0200-1

Benford, S., Greenhalgh, C., Rodden, T., & Pycock, J. (2001). Collaborative virtual environments. *Communications of the ACM*, *44*(7), 79–85. https://doi.org/10.1145/379300.379322

Bergström, J., Knibbe, J., Pohl, H., & Hornbæk, K. (2022). Sense of agency and user experience: Is there a link? *ACM Transactions on Computer-Human Interaction*, *29*(4), 1–22. https://doi.org/10.1145/3490493

Brandis, R., & Bozkurt, C. M. (2021). Player agency in audience gaming. In B. Beil, G. S. Freyermuth, & H. C. Schmidt (Eds.), *Paratextualizing games* (pp. 165–180). Transcript Verlag.

Carillo, K., & Okoli, C. (2011). Generating quality open content: A functional group perspective based on the time, interaction, and performance theory. *Information & Management*, *48*(6), 208–219. https://doi.org/10.1016/j.im.2011.04.004

Cerratto-Pargman, T., Rossitto, C., & Barkhuus, L. (2014). *Understanding audience participation in an interactive theater performance* [Paper presentation]. In Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational (pp. 608–617). https://doi.org/10.1145/2639189.2641213

Chen, C. C., & Lin, Y. C. (2018). What drives live-stream usage intention? The perspectives of flow, entertainment, social interaction, and endorsement. *Telematics and Informatics*, *35*(1), 293–303. https://doi.org/10.1016/j.tele.2017.12.003

Chen, L., Liang, H.-N., Lu, F., Wang, J., Chen, W., & Yue, Y. (2021). Effect of collaboration mode and position arrangement on immersive analytics tasks in virtual reality: A pilot study. *Applied Sciences*, *11*(21), 10473. https://doi.org/10.3390/app112110473

Chen, L., Liu, Y., Li, Y., Yu, L., Gao, B., Caon, M., Yue, Y., & Liang, H.-N. (2021). *Effect of visual cues on pointing tasks in co-located augmented reality collaboration* [Paper presentation]. In Proceedings of the 2021 Acm Symposium on Spatial User Interaction. https://doi.org/10.1145/3485279.3485297

Deng, J., Cuadrado, F., Tyson, G., & Uhlig, S. (2015). *Behind the game: Exploring the twitch streaming platform* [Paper presentation]. 2015 International Workshop on Network and Systems Support for Games (NetGames) (pp. 1–6). https://doi.org/10.1109/NetGames.2015.7382994

Fan, S., Liu, J., & Zhu, T. (2018). *Role of danmu function in user experience and engagement: A double-edged sword* [Paper presentation]. Wuhan International Conference on e-Business.

Ford, C., Gardner, D., Horgan, L. E., Liu, C., Tsaasan, a m., Nardi, B., & Rickman, J. (2017). *Chat speed op pogchamp: Practices of coherence in massive twitch chat* [Paper presentation]. In Proceedings of the 2017 Chi Conference Extended Abstracts on Human Factors in Computing Systems, (p. 858–871). Association for Computing Machinery. https://doi.org/10.1145/3027063.3052765

Getzlaf, B., Melrose, S., Moore, S., Ewing, H. L., Fedorchuk, J., & Troute-Wood, T. (2012). Online interest groups: Virtual gathering spaces to promote graduate student interaction. *International Journal of Online Pedagogy and Course Design*, *2*(4), 63–76. https://doi.org/10.4018/ijopcd.2012100105

Glickman, S., McKenzie, N., Seering, J., Moeller, R., & Hammer, J. (2018). *Design challenges for livestreamed audience participation games* [Paper presentation]. Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play (p. 187–199). Association for Computing Machinery. https://doi.org/10.1145/3242671.3242708

Haimson, O. L., & Tang, J. C. (2017). *What makes live events engaging on facebook live, periscope, and snapchat* [Paper presentation]. Proceedings of the 2017 Chi Conference on Human Factors in Computing Systems, In (p. 48–60). Association for Computing Machinery. https://doi.org/10.1145/3025453.3025642

Hart, S. G., & Staveland, L. E. (1988). Development of nasa-tlx (task load index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (Vol. 52, p. 139–183).

Hassenzahl, M. (2004). The interplay of beauty, goodness, and usability in interactive products. *Human-Computer Interaction*, *19*(4), 319–349. https://doi.org/10.1207/s15327051hci1904_2

He, D., Yao, Z., Tang, P., & Ma, Y. (2023). Impacts of different interactions on viewers' sense of virtual community: An empirical study of live streaming platform. *Behaviour & Information Technology*, *42*(7), 940–960. https://doi.org/10.1080/0144929X.2022.2053884

Hilvert-Bruce, Z., Neill, J. T., Sjöblom, M., & Hamari, J. (2018). Social motivations of live-streaming viewer engagement on twitch. *Computers in Human Behavior*, *84*, 58–67. https://doi.org/10.1016/j.chb.2018.02.013

Hu, M., Zhang, M., & Wang, Y. (2017). Why do audiences choose to keep watching on live video streaming platforms? An explanation of dual identification framework. *Computers in Human Behavior*, *75*(4), 594–606. https://doi.org/10.1016/j.chb.2017.06.006

Kim, J. (2013). Influence of group size on students' participation in online discussion forums. *Computers & Education*, *62*, 123–129. https://doi.org/10.1016/j.compedu.2012.10.025

King, D. L., Delfabbro, P. H., Billieux, J., & Potenza, M. N. (2020). Problematic online gaming and the covid-19 pandemic. *Journal of Behavioral Addictions*, *9*(2), 184–186. https://doi.org/10.1556/2006.2020.00016

Lasecki, W. S., Murray, K. I., White, S., Miller, R. C., & Bigham, J. P. (2011). *Real-time crowd control of existing interfaces* [Paper presentation]. Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology (p. 23–32). Association for Computing Machinery. https://doi.org/10.1145/2047196.2047200

Lee, L.-H., Braud, T., Zhou, P., Wang, L., Xu, D., Lin, Z., & Hui, P. (2021). All one needs to know about metaverse: A complete survey on technological singularity, virtual ecosystem, and research agenda. *arXiv Preprint arXiv:2110.05352*.

Lesniak, K. K., & Maistro, M. (2022). *Crowdsourcing controller - utilizing reliable agents in a multiplayer game* [Paper presentation]. 2022 IEEE Conference on Games (Cog) (p. 64–71). IEEE. https://doi.org/10.1109/CoG51982.2022.9893597

Lessel, P., Vielhauer, A., & Krüger, A. (2017). *Expanding video game live-streams with enhanced communication channels: A case study* [Paper presentation]. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (p. 1571–1576). Association for Computing Machinery. https://doi.org/10.1145/3025453.3025708

Lewis, J. R. (2018). Measuring perceived usability: The CSUQ, SUS, and UMUX. *International Journal of Human–Computer Interaction*, *34*(12), 1148–1156. https://doi.org/10.1080/10447318.2017.1418805

Li, Y., & Guo, Y. (2021). Virtual gifting and danmaku: What motivates people to interact in game live streaming? *Telematics and Informatics*, *62*(3), 101624. https://doi.org/10.1016/j.tele.2021.101624

Li, Y., Wang, C., & Liu, J. (2020). A systematic review of literature on user behavior in video game live streaming. *International Journal of Environmental Research and Public Health*, *17*(9), 3328. https://doi.org/10.3390/ijerph17093328

Liang, H.-N., Lu, F., Shi, Y., Nanjappan, V., & Papangelis, K. (2019). Evaluating the effects of collaboration and competition in navigation tasks and spatial knowledge acquisition within virtual reality environments. *Future Generation Computer Systems*, *95*(2), 855–866. https://doi.org/10.1016/j.future.2018.02.029

Liang, H.-N., Parsons, P. C., Wu, H.-C., & Sedig, K. (2010). An exploratory study of interactivity in visualization tools: 'flow' of interaction. *Journal of Interactive Learning Research*, *21*(1), 5–45.

Liang, H.-N., & Sedig, K. (2010). Can interactive visualization tools engage and support pre-university students in exploring non-trivial mathematical concepts? *Computers & Education*, *54*(4), 972–991. https://doi.org/10.1016/j.compedu.2009.10.001

Liu, H., Tan, K. H., & Pawar, K. (2022). Predicting viewer gifting behavior in sports live streaming platforms: The impact of viewer perception and satisfaction. *Journal of Business Research*, *144*(11), 599–613. https://doi.org/10.1016/j.jbusres.2022.02.045

Lombard, M., Weinstein, L., & Ditton, T. (2011). *Measuring telepresence: The validity of the temple presence inventory (tpi) in a gaming context* [Paper presentation]. Ispr 2011: The International Society for Presence Research Annual Conference.

Longo, M. R., & Haggard, P. (2009). Sense of agency primes manual motor responses. *Perception*, *38*(1), 69–78. https://doi.org/10.1068/p6045

Lu, Z., Heo, S., & Wigdor, D. J. (2018). Streamwiki: Enabling viewers of knowledge sharing live streams to collaboratively generate archival documentation for effective in-stream and post hoc learning. *Proceedings of the ACM on Human-Computer Interaction*, *2*(CSCW), 1–26. https://doi.org/10.1145/3274381

Lytle, C., Ramsey, P., Yeo, J., Dressen, T., Kang, D. h., Harger, B. B., & Hammer, J. (2020). *Toward live streamed improvisational game experiences* [Paper presentation]. In Proceedings of the Annual Symposium on Computer-Human Interaction in Play, (p. 148–159).

Association for Computing Machinery. https://doi.org/10.1145/3410404.3414226

Mallory, M. (2014). Community-based play in twitch plays pokémon. *Well Played*, *3*(2), 25–43.

Monteiro, D., Ma, T., Li, Y., Pan, Z., & Liang, H.-N. (2022). Cross-cultural factors influencing the adoption of virtual reality for practical learning. *Universal Access in the Information Society*, 1–14. https://doi.org/10.1007/s10209-022-00947-y

Oldenburg, R. (1999). The great good place: Cafes, coffee shops, bookstores, bars, hair salons, and other hangouts at the heart of a community. Distributed by Publishers Group West.

Pascal, L., Michael, M., Christian, W., & Kruger, A. (2017). *Let's play my way: Investigating audience influence in user-generated gaming live-streams* [Paper presentation]. In Proceedings of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video (p. 51–63). Association for Computing Machinery. https://doi.org/10.1145/3077548.3077556

Pires, K., & Simon, G. (2015). *Youtube live and twitch: A tour of user-generated live streaming systems* [Paper presentation]. In Proceedings of the 6th ACM Multimedia Systems Conference (p. 225–230). Association for Computing Machinery. https://doi.org/10.1145/2713168.2713195

Poirier-Poulin, S. (2020). Watch me play: Twitch and the rise of game live streaming. *New Media & Society*, *22*(8), 1505–1507. https://doi.org/10.1177/1461444820939317

Ramirez, D., Saucerman, J., & Dietmeier, J. (2014). Twitch plays pokemon: A case study in big g games. *Proceedings of Digra* (pp. 3–6).

Sakamoto, M., Nakajima, T., & Akioka, S. (2017). Gamifying collective human behavior with gameful digital rhetoric. *Multimedia Tools and Applications*, *76*(10), 12539–12581. https://doi.org/10.1007/s11042-016-3665-y

Sedig, K., & Parsons, P. C. (2016). *Design of visualizations for human-information interaction: A pattern-based framework*. Springer Cham.

Seering, J., Savage, S., Eagle, M., Churchin, J., Moeller, R., Bigham, J. P., & Hammer, J. (2017). *Audience participation games: Blurring the line between player and spectator* [Paper presentation]. In Proceedings of the 2017 Conference on Designing Interactive Systems (p. 429–440). Association for Computing Machinery. https://doi.org/10.1145/3064663.3064732

Shaw, R.-S. (2013). The relationships among group size, participation, and performance of programming language learning supported with online forums. *Computers & Education*, *62*, 196–207. https://doi.org/10.1016/j.compedu.2012.11.001

Shiau, W.-L., & Huang, L.-C. (2023). Scale development for analyzing the fit of real and virtual world integration: An example of pokémon go. *Information Technology & People*, *36*(2), 500–531. https://doi.org/10.1108/ITP-11-2020-0793

Slater, M., Usoh, M., & Steed, A. (1994). Depth of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, *3*(2), 130–144. https://doi.org/10.1162/pres.1994.3.2.130

Statista. (2023). *Average number of monthly active users of Bilibili Inc. from Q2 2019 to Q2 2023*. Retrieved January 29, 2024, from https://www.statista.com/statistics/1109108/bilibili-average-monthly-active-users/

Striner, A., Azad, S., & Martens, C. (2019). *A spectrum of audience interactivity for entertainment domains* (Vol. 11869). LNCS.

Striner, A., Webb, A. M., Hammer, J., & Cook, A. (2021). *Mapping design spaces for audience participation in game live streaming* [Paper presentation]. Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery. https://doi.org/10.1145/3411764.3445511

StudioBean. (2016). *Choice chamber*. http://choicechamber.com/sub

Synofzik, M., Vosgerau, G., & Newen, A. (2008). Beyond the comparator model: A multifactorial two-step account of agency. *Consciousness and Cognition*, *17*(1), 219–239. https://doi.org/10.1016/j.concog.2007.03.010

TwitchInstallsArchLinux. (2024). *Twitch installs arch linux*. Retrieved January 25, 2024, from https://www.twitch.tv/twitchinstallsarchlinux

TwitchPlaysDark. (2024). *Twitch plays dark souls*. Retrieved January 25, 2024, from https://www.twitch.tv/twitchplaysdark

TwitchTracker. (2024). *Twitch statistics & charts*. Retrieved January 25, 2024, from https://twitchtracker.com/statistics

Wang, X., Tian, Y., Lan, R., Yang, W., & Zhang, X. (2019). Beyond the watching: Understanding viewer interactions in crowdsourced live video broadcasting services. *IEEE Transactions on Circuits and Systems for Video Technology*, *29*(11), 3454–3468. https://doi.org/10.1109/TCSVT.2018.2877694

Xu, W., Zheng, R., Monteiro, D., Nanjappan, V., Wang, Y., & Liang, H.-N. (2024). *Exploring the effect of display type on co-located multiple player gameplay performance, immersion, social presence, and behavior patterns* [Paper presentation]. Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Grapp (p. 159–169). https://doi.org/10.5220/0012469000003660

Zhang, H., Yang, Y., & Zhao, J. (2023). Does game-irrelevant chatting stimulate high-value gifting in live streaming? A session-level perspective. *Computers in Human Behavior*, *138*(8), 107467. https://doi.org/10.1016/j.chb.2022.107467

Zhou, J., Zhou, J., Ding, Y., & Wang, H. (2019). The magic of danmaku: A social interaction perspective of gift sending on live streaming platforms. *Electronic Commerce Research and Applications*, *34*, 100815. https://doi.org/10.1016/j.elerap.2018.11.002

About the authors

Yue Li is an Assistant Professor in the Department of Computing at Xi'an Jiaotong-Liverpool University. Her research interest is in the field of human-computer interaction, with particular emphasis on the design, evaluation, and application of virtual and augmented reality technologies in cultural heritage and education.

Teng Ma is an Assistant Professor in the Department of Computing at Xi'an Jiaotong Liverpool University. His research interests focus on exploring the impact of technologies in information dissemination, social computing, and business analytics.

Ziming Li is a first-year Ph.D. student at Xi'an Jiaotong-Liverpool University. His research focuses on Virtual Reality (VR) and its applications, especially for collaborative and teleoperation tasks.

Hai-Ning Liang is an Associate Professor in the Computational Media and Arts Thrust at the Hong Kong University of Science and Technology (Guangzhou). His main research interests fall in the area of human-computer interaction, focusing on designing novel techniques and applications for virtual/augmented/mixed reality, visualization, and gaming technologies.