2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)

Towards Cross-Reality Interaction and Collaboration: A Comparative Study of Object Selection and Manipulation in Reality and Virtuality

Shuhao Zhang [1], Yue Li [1,*], Ka Lok Man [1], Yong Yue [1], and Jeremy Smith[2]
[1] School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou, China
[2] Department of Electrical Engineering and Electronics, University of Liverpool, United Kingdom

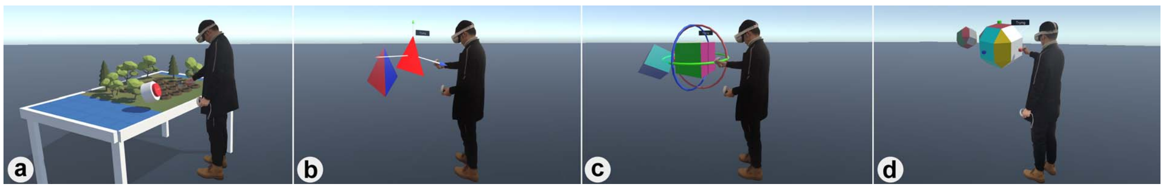


Figure 1: Demonstration of a VR user completing (a) selection and (b-d) transform manipulation tasks (move, rotate, and scale) .

Abstract

Cross-Reality (CR) is an important topic for the research of multiuser collaborative systems. It allows users to participate in the reality-virtuality continuum and select appropriate interactive systems to work with, such as Virtual Reality Head-Mounted Displays (VR HMDs). However, there is limited work showing how interaction in VR differs from the more commonly used Personal Computers (PCs) and tablet devices in terms of object selection and manipulation. In this paper, we present a comparative study that investigated how users perform and perceive workload on 3D object selection and manipulation tasks using different devices (e.g. PC, tablet, and VR). We recorded the time and accuracy as objective task performance measures, and users' self-reported workload as a subjective measure. Our results revealed that unlike the biased performances of PC and tablet, VR has a balanced performance and great potentials in complex tasks.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Virtual reality; Human-centered computing—Human computer interaction (HCI)—HCI design and evaluation methods—User studies

1 Introduction

Cross-Reality (CR) systems aim to enable interactions across multiple technologies and allow users to collaborate through different devices. The development of immersive technology such as Augmented Reality (AR) and Virtual Reality (VR) has brought more possibilities for the development of cross-reality systems. Based on the Reality-Virtuality continuum [18], it is envisioned that future CR systems to allow a single user to have a smooth transition between systems using different degrees of virtuality, and multiple users to collaborate using different systems with different degrees of virtuality [26]. For example, it allows users in VR to interact with co-located [7] or remote [28] spectators in the real world with [7, 8, 28] or without [12, 32] other devices' help. In addition, this method can also effectively reduce the isolation of HMD users and the exclusion of bystanders (i.e., non-HMD users) [8, 25].

Some studies have tried to explore the combined use of different interactive devices. For example, Kobayashi et al. [13] studied different task performance for 2D manipulation tasks using tablets with different screen sizes and found that it is a suitable device for the elderly, but designers need to avoid high-precision tap tasks when the screen size is small. Li et al. [15] studied the acceptance of hybrid use of VR HMD and mobile AR, and found the significant impact of social influence on users' behavioural intention. However, these studies did not compare the difference between interactive devices used in reality and virtuality. The differences in users' performance and subjective experience in reality and virtuality are important parts of future cross-reality systems, but current work in this area is limited. In such situations, clarifying the characteristics of interactive devices (e.g. PC, tablets, VR) is meaningful and will contribute to the research community.

In this study, we designed a comparative study to explore how users' performance and perceived workload vary in different 3D object interaction tasks (selection and manipulation) when using different devices (PC, tablet, and VR). The main contributions are three-fold. First, we provided an empirical evaluation of users' task performances and perceived workload using PC, tablets, and VR when doing selection and manipulation tasks. Second, our study design took into account three levels of task difficulty in selection tasks and three transform operations (moving, rotating, and scaling) in manipulation tasks, showing results that are likely to be generalisable and applicable to different contexts. Third, we discussed the features of each device, which has led to useful findings and design implications for future cross-reality interactive systems in various fields such as design, education, and games.

*Corresponding author: yue.li@xjtlu.edu.cn

2 Related Work

2.1 Interactive Devices in Reality and Virtuality

In reality, Personal Computers (PCs) and mobile devices are two of the most commonly used interactive devices. Users' interactions with PCs are often mediated by a set of mouse and keyboard, and users interact with mobile devices through touchscreen control. Previous work [31] has shown that the PC control based on a mouse and a keyboard requires a high demand for users' hand-eye coordination, consequently, more cognitive effort. Compared with PCs, mobile device control through touchscreens is more direct. Users will receive instant feedback from the device when they touch different portions of the screen with a finger or stylus. For novice users, such direct interaction is easy to learn and operate [11, 21]. However, researchers have also found that there is often a deviation between users' expected contact point and the actual finger position. Such deviations will affect users' task performance.

In virtuality, VR Head-Mounted Display (HMD) is one of the most popular devices and has been applied in various fields, such as education [19], training [30], gaming [2], and culture heritage [29]. When wearing an HMD, users could interact with the virtual en-

979-8-3503-4839-2/23/$31.00 ©2023 IEEE
DOI 10.1109/VRW58643.2023.00075

vironments (VEs) using hand-held controllers. In some research prototypes, devices such as VR gloves [5], eye-tracking [17], depth cameras [6], and motion capture devices [10] were also used for VR controls. Fahmi et al. [3] compared users' acceptability, satisfaction, system learnability, and haptic feedback among three different control devices for anatomical learning: VIVE controller, Leap motion, and Senso glove. The authors have found that except for the greater haptic feedback in Senso glove, VIVE controller was significantly better than Leap motion and Senso glove in other aspects. Controller-based interaction is still the most common approach that has been adopted in commercial VR HMDs (such as HTC VIVE and Meta Quest) and VR applications [3, 17].

Previous studies have provided insights into the features of interactive devices in reality and virtuality, and pointed out that they vary in acceptability, learnability, usability, and technology affordance. Despite the understanding of the differences, how the device features and the trade-offs affect user interactions in CR systems is underexplored. Future development of CR systems and applications is likely to involve transitions between devices used in reality and virtuality, and to have users using different devices collaborate across the reality-virtuality continuum. However, the relationship between the degree of reality/virtuality and users' performance and workload remains to be explored.

2.2 Cross-Reality Interaction and Collaboration

Cross-reality systems offer different levels of physicality and virtuality to users and enable them to move between the reality-virtuality continuum in a seamless way. In recent years, the popularity of VR and AR has brought some explorations in CR interaction and collaboration as well as prototypes that demonstrate potential use cases. Two types of work can be seen in the literature.

Some studies have tried to visualise the VEs in the real world to engage users in reality. For example, FrontFace [1] and FaceDisplay [8] allow non-HMD users in the real world to observe the VEs and interact with the VR HMD user by the installed touchscreen display in front of and around the VR HMD. ShareVR [7] attempted to engage co-located non-HMD users through floor projections and hand-held controllers, and found it to have increased users' enjoyment, presence, and social interaction. HVAR [16] combined VR and AR for a shared experience, and showed that AR can be used for including audiences in scenarios that support the spectator experience and for complementing and enriching VR in social contexts. Sra et al. [27] designed a visualisation method for a multiplayer shooting game using cardboard boxes and servos. They made physical replicas of target blocks in the VE, so that the physical box collapses when the virtual block collapses. Different from digital displays, this study reflected the changes in the VE to the real world through physical mappings.

Another type of work aims to leverage the features of reality and virtuality for collaborative work. For example, Radu et al. [23] presented a system that allows VR HMD users, AR HMD users, and PC users to interact with physical objects and overlays together. Similar multi-device systems can be seen in education research [19, 20], where the teacher wearing a VR HMD and the students can join the virtual learning environment using other devices. However, there were no formal user studies evaluating the effectiveness of these systems. Mini-Me [22] studied the collaboration between VR and AR users and showed that an adaptive avatar representing the VR user's gaze direction and body gestures could enhance users' collaboration. The authors evaluated users' coordinated actions in object placement tasks, which do not require precise control of object transforms.

Recent work has demonstrated some interesting use cases of early CR systems, such as learning and education [19, 20, 23], games [7], and urban planning [22]. However, many of them focused on the design aspect and lack evaluation studies; those with evaluations were mainly based on object placement tasks. More complex tasks requiring precise control of object transforms are likely to be seen in future CR systems for fields such as architectural design, automotive engineering, and surgery training. Thus, an investigation of fundamental object selection and manipulation tasks, including the operations of object transforms (i.e. position, rotation, and scale) is needed for widely applicable CR system research.

3 METHODOLOGY

3.1 System Development

Our proof-of-concept prototype was built using a computer with Intel Core i7-9750H CPU @ 2.60GHz, 8GB RAM, NVIDIA GeForce GTX 1650 graphics card with 4GB RAM. The systems were built using the Unity engine (version 2021.3.7) and two packages: VR Interaction Framework[1] and Runtime Transform Gizmos[2], which are available on the Unity Assort Store. We used 3D Studio Max 2016 and Rhino 7.0 to build the 3D models and set up virtual scenes.

As for experimental facilities, we adopted three devices for the three experimental conditions: a PC, a tablet, and a VR HMD. The same laptop used for the system development with a 15.6 inch screen display (1920×1080 resolution, 60 Hz refresh rate) was used for the PC condition; a Mi 4 Plus tablet with a 10.1 inch display (1920×1080 resolution, 30 Hz refresh rate) was used for the tablet (TB) condition; and an Oculus Quest 2 (1920×1832 resolution for each eye, 72 Hz refresh rate) was used for the VR condition.

We set up some C# scripts to capture objective user behaviour data. The program prints all records on the console, which were exported as CSV files of each session for further analysis. Table 1 shows some example lines of user behaviour data.

Table 1: Example CSV records for WHAC-A-MOLE and OVERLAPPING.

Scene	Real time	Game time	Object name	State
Mole_L2	17:15:17	110.71	mole_05	Appear
Mole_L2	17:15:17	111.03	mole_15	Appear
Mole_L2	17:15:17	111.17	mole_05	Success
Mole_L2	17:15:18	111.28	Ground	Fail
...	...	...	...	...
M	17:19:02	341.54	Hexahedron01	Start
M	17:19:20	353.86	Hexahedron01	Success

3.2 Study Design

The study was conducted using a repeated measures factorial design with two independent variables: Device (PC, TB, and VR) and Task (Selection and Manipulation). Selection and manipulation are two typical 3D object interaction tasks [14]. Thus, we include these two common 3D object interaction tasks in our study design, and designed two interactive systems (WHAC-A-MOLE and OVERLAPPING) to explore the selection and manipulation of 3D objects.

3.2.1 WHAC-A-MOLE

As a classic game, WHAC-A-MOLE follows a simple principle: hit and destroy moles as they appear. This requires users to perform selection tasks repeatedly. In order to obtain an in-depth understanding of user performance with 3D object selection using three different devices, we set up three difficulty levels (L1, L2, L3) based on Fitt's law. As Fitt's law indicates that the amount of time required for a selection task is affected by the distance to the target and the size of the target, we placed different numbers of moles within the same play area to simulate different levels of task difficulty.

Task Difficulty. We set up three sessions of different difficulty levels. In L1, the simplest session, nine (3×3) holes were placed in

[1]https://assetstore.unity.com/packages/templates/systems/vr-interaction-framework-161066.

[2]https://assetstore.unity.com/packages/tools/modeling/runtime-editor-64806.

the game area (see Figure 2a), with standard-size moles popping up; In L2, users faced with sixteen (4×4) holes in the game area, each in a 0.8 standard size (see Figure 2b); In L3, the most difficult session, twenty-five (5×5) holes were placed in the game area (see Figure 2c) and each mole was in a 0.6 standard size. Each session lasted one minute, during which the moles continued to pop up, and the refresh rate increased linearly from three moles per group to seven moles per group. All moles that appeared stayed for one second.

Figure 2: Screenshots of PC interfaces demonstrating the selection task setup based on WHAC-A-MOLE, including three sessions of different difficulty levels: (a) L1: 3×3 blocks, (a) L2: 4×4 blocks; (c) L3: 5×5 blocks.

Control. For PC, users need to move the mouse and click to hit the moles. For tablet, users need to hit the moles by tapping their fingers on the screen as they appeared. When using VR, users need to hold a hammer to hit the moles (see Figure 1a).

3.2.2 OVERLAPPING

OVERLAPPING aims to explore the effect of the devices on the manipulation of virtual objects. This task requires users to change the position and rotation translations and the size of 3D objects through moving, rotating, and scaling operations and finally make it overlap with the target object. To make the control consistent across the three devices for a valid comparison, the manipulation of 3D objects was achieved based on operations on three axes (x, y, and z), including VR. Although this may not be the most common and efficient method of object manipulation in VR [24], this is a common technique used in both PC and tablet. Besides, design works such as 3D modelling and building constructions often require precise control and have rigorous requirements of positions and scales. In this case, manipulations of the three transforms offer more room for control than direct manipulations. Thus, we implemented and evaluated this technique in VR.

Transform Operations. In order to obtain an in-depth understanding of user performance with 3D object manipulation, we investigated user performance for each transform operation through four sessions: movement (M), rotation (R), scaling (S), and a hybrid control that integrates the three transform operations (MRS). Figure 3 provides an illustration of the four sessions.

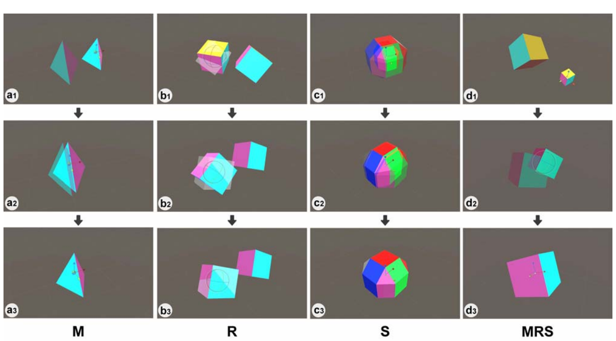


Figure 3: Screenshots of PC interfaces demonstrating of the OVERLAPPING setup for manipulation tasks. (a1-a3) Movement, (b1-b3) Rotation, (c1-c3) Scaling, and (d1-d3) MRS hybrid tasks. Three objects were used: (a1-a3) tetrahedron, (b1-b3) hexahedron, and (c1-c3) octadecahedron.

Trials on Three Objects. We set up three objects for each session: tetrahedron, hexahedron, and octadecahedron (see Figure 3). Target points were set up at each vertex of the object, in which case there were four for tetrahedron, eight for hexahedron, and sixteen for octadecahedron. When the object collides with all target points, the task is completed, and the participant is informed of the task status by checking if the icon above the object shows "*Trying*" or "*Success*".

Control. The manipulation control of 3D objects using PC is similar to most 3D modelling and development software interfaces, such as Unity. Users need to manipulate objects by clicking and dragging the control axis (see Figure 3). To switch control commands, users can either use keyboard commands (W: move, E: rotate, R: scale) or click on UI icons using the mouse. For tablets, users need to control objects by dragging the control axis, and tapping their fingers on the UI icons to switch commands. For VR, users need to choose the control axis to realise object control when they pressing the Grip button using the index finger. Similar to the PC control switch, users can press the A, B, and X buttons on the controllers to activate move, rotate, and scale controls.

3.3 Procedure and Tasks

We conducted a within-subjects study that took place in a 2m×3m space in a university lab. After a brief introduction and collecting participants' consent, participants were asked to familiarise with the devices and adjust them to their most comfortable state, including the sensitivity of the mouse and the fit and focus of the VR HMD. Participants were required to finish a training session for each task (WHAC-A-MOLE and OVERLAPPING) to get familiar with the systems and controls. After the tutorial, there were six experimental sessions (3 Devices × 2 Task Types). A Latin square design was applied to avoid the influence of experimental order on the results. Participants were asked to fill in a questionnaire after each session to evaluate their workload, after which they were encouraged to rest fully and inform the researcher when they were ready for the next experimental session. We concluded the experiment with a debriefing session and an open discussion. The experiment lasted ∼110 minutes on average, including a tutorial (∼10 minutes), selection tasks (∼15 minutes), manipulation tasks (∼60 minutes), rest time (∼15 minutes), and an open discussion (∼10 minutes). This study has received ethics approval from the University Ethics Committee at the Xi'an Jiaotong-Liverpool University.

3.4 Measures

The aim of our work is to evaluate the differences in the task performance and workload between different devices and tasks. Therefore, we used the NASA Task Load Index (NASA-TLX) [9] to obtain self-reported workload measure. Users' task performance in each task was collected and calculated using instrumented automated data collection through C# scripts in Unity.

Workload. The NASA-TLX questionnaire consists of six questions to analyse users' workload. Each question assesses user's feelings on one of the six dimensions: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. Questions were rated on a scale between 0 and 100. Users were asked to report on the NASA-TLX questionnaire after each sub-task. Therefore, a total of 21 (3 Devices × 7 Task sessions) responses were collected from each user.

Task Performance (Accuracy) in WHAC-A-MOLE. We recorded all click events in the selection tasks. We labelled a click as "*Success*" when a mole was hit, and "*Fail*" otherwise. For each session, we counted the number of successful trials (S) and the number of failed trials (F). The accuracy (A) was thus calculated by the formula: $A = \frac{S}{S+F}$.

Task Performance (Time) in OVERLAPPING. In the manipulation tasks, we used time to evaluate the user's task performance. The timer started when a user entered the task environment, and stopped when the task status changed to "*Success*", i.e., the task object overlapped with the target object.

Data analysis was performed using IBM SPSS Statistics 26. We performed repeated measures ANOVA to analyse the effects of device and task on performance and workload. Greenhouse-Geisser correction was applied when the collected data does not satisfy the sphericity test assumption, and Bonferroni adjustment was applied for post hoc tests to avoid inflated Type I error. We report Bonferroni-adjusted p-values, i.e., multiplying the observed (uncorrected) p value by the number of comparisons made, which is compared against the threshold value of .05.

4 RESULTS

We had 12 participants (4 female, 8 male) who voluntarily signed up for the study, with an average age of 23.67 (SD= 3.34). Participants were asked to rate their usage frequencies and familiarity with the three types of devices, and their abilities in drawing and 3D modelling. Table 2 shows the usage frequencies of difference devices. On a 5-point Likert scale, participants' familiarity of different devices from the highest to the lowest was PC (4.75 ± .452), TB (4.17 ± .937) and VR (4.00 ± .739). We also asked participants to rate on their 3D modelling abilities (2.42 ± .793).

Table 2: Usage frequencies for each device.

	PC	TB	VR
Never	0	0	0
Less than 1 time per month	0	2	0
1-3 times per month	0	2	4
1-3 times per week	0	4	3
4-7 times per week	12	4	5

4.1 Selection Task Evaluation: WHAC-A-MOLE

4.1.1 Selection Accuracy

A two-way repeated measures ANOVA was conducted to examine the effects of device and task on selection accuracy. There was a statistically significant interaction between the effects of device and task in selection accuracy, $F(4, 44) = 24.360$, $p < .001$. Further analysis showed that both device ($F(2, 22) = 15.308$, $p < .001$) and task ($F(2, 22) = 57.228$, $p < .001$) have a statistically significant effect on selection accuracy.

Device. Simple main effects analysis revealed that TB showed significantly higher selection accuracy than PC in L1 (p = .005), L2 (p = .001), and L3 (p <.001). VR showed higher selection accuracy than PC in L1 (p = .906) and L2 (p = .105), but the difference was not significant. When the difficulty level increases to L3, the accuracy of VR became significantly higher than PC (p <.001). Overall, there were significant differences between PC-TB (p <.001) and PC-VR (p = .015). Selection accuracy was significantly lower in PC, compared to VR and TB (see Figure 4).

Task Difficulty. Simple main effects analysis showed that there was no significant difference in L1-L2 when using PC (p = .981) or TB (p = 1.000), but significant differences between L1-L3 (p <.001) and L2-L3 (p <.001). When using VR, the differences in L1-L2 (p = .367), L1-L3 (p = .305), and L2-L3 (p = 1.000) were insignificant. Overall, there was no significant difference between L1-L2 (p = 1.000), but a significant difference between L1-L3 (p <.001) and L2-L3 (p <.001). Accuracy in L3 was significantly lower than L1 and L2 (see Figure 4).

4.1.2 Workload

A two-way repeated measures ANOVA showed no significant interaction between the effects of device and task on workload, $F(4, 44) = 2.170$, $p = .088$. Further analysis showed that both device ($F(2, 22) = 8.350$, $p = .002$) and task ($F(2, 22) = 24.671$, $p < .001$) have a significant effect on workload. Post hoc analysis with a Bonferroni adjustment revealed significant differences between PC-TB (p = .018) and TB-VR (p = .011). Workload was significantly lower with TB (23.694 ± 2.297), compared to PC (39.074 ± 4.608) and VR (41.366 ± 4.312) (see Figure 5). For different levels, there were significant differences between L1-L2 (p = .037), L1-L3 (p = .001) and L2-L3 (p <.001). Workload from the highest to the lowest was L3, L2, and L1 (see Figure 5).

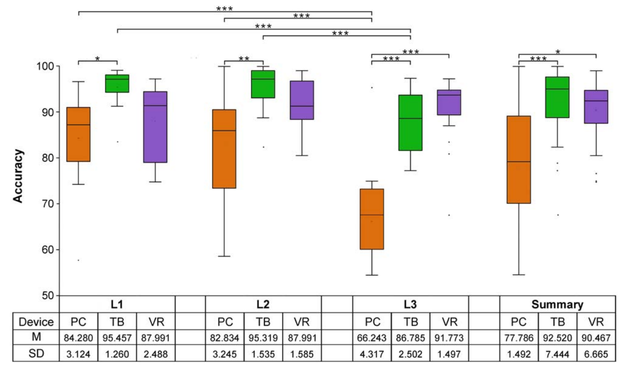


Figure 4: Accuracy results in WHAC-A-MOLE selection tasks.

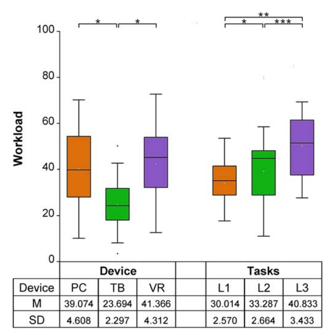


Figure 5: Workload results in WHAC-A-MOLE selection tasks based on Device and Task.

4.2 Manipulation Task Evaluation: OVERLAPPING

4.2.1 Manipulation Time

Figure 6 showed the results of users' manipulation time in the four sessions: (M) Movement, (R) Rotation, (S) Scaling, and (MRS) hybrid tasks that combined three transform operations.

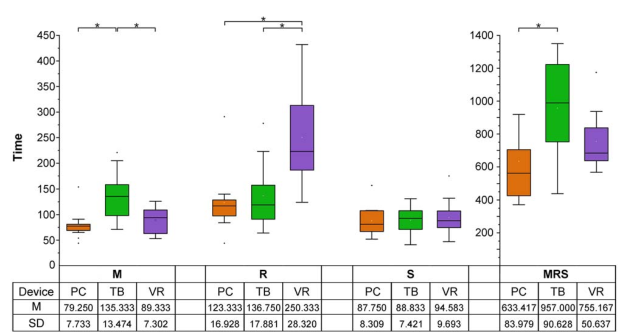


Figure 6: Time results for manipulation tasks in OVERLAPPING: (M) Movement, (R) Rotation, (S) Scaling, and (MRS) hybrid tasks.

Movement. A repeated measures ANOVA determined that manipulation time differed significantly between devices in Movement tasks, $F(2, 22) = 16.786$, $p < .001$. Post hoc analysis with a Bonferroni adjustment revealed that the difference in time was significant between PC-TB (p = .002) and TB-VR (p = .005), but insignificant between PC-VR (p = .591).

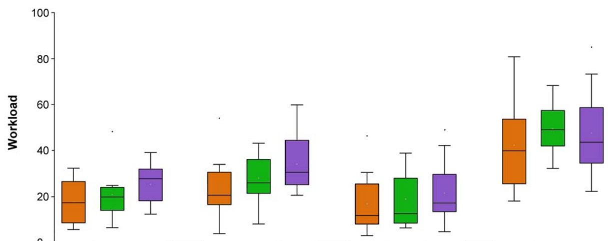
Rotation. A repeated measures ANOVA determined that manipulation time differed significantly between devices in Rotation tasks, $F(2, 22) = 9.537$, $p = .001$. Post hoc analysis with a Bonferroni adjustment revealed significant difference in time between PC-VR ($p = .010$) and TB-VR ($p = .022$), but not PC-TB ($p = 1.000$).

Scaling. A repeated measures ANOVA determined that manipulation time did not differ significantly between devices in Scaling tasks, $F(2, 22) = .247$, $p = .783$.

MRS. A repeated measures ANOVA determined that manipulation time differed significantly between devices in MRS tasks, $F(2, 22) = 5.725$, $p = .010$. Post hoc analysis with a Bonferroni adjustment revealed that time was significantly different between PC-TB ($p = .040$), but insignificant between TB-VR ($p = .065$) and PC-VR ($p = .762$). Ranking the time from the longest to the shortest was TB (957.00 ± 90.63), VR (755.167 ± 50.64) and PC (633.42 ± 83.98).

4.2.2 Workload

Figure 7 shows the analysis results of users' workload. Repeated measures ANOVAs showed that both device ($F(2, 22) = 3.480$, $p = .049$) and task ($F(1.766, 19.423) = 40.059$, $p < .001$) have a significant effect on workload. Post hoc analysis with a Bonferroni adjustment revealed no significant differences in pair-wise comparisons. The differences in workload were insignificant in Movement, Rotation, Scaling, or MRS hybrid tasks using three devices.



	M			R			S			MRS		
Device	PC	TB	VR	PC	TB	VR	PC	TB	VR	PC	TB	VR
M	17.569	20.972	24.486	23.625	28.264	34.306	17.153	19.167	21.833	42.431	49.583	47.722
SD	2.853	2.983	2.461	3.726	3.051	3.523	3.729	3.497	3.929	5.732	3.261	5.318

Figure 7: Workload results of the manipulation tasks in OVERLAPPING: (M) Movement, (R) Rotation, (S) Scaling, and (MRS) hybrid tasks.

4.2.3 Additional Findings

We report our additional findings on the success rate in manipulation tasks. For the manipulation of tetrahedron, hexahedron, and octadecahedron, the success rate were 86.11%, 94.5%, and 97.23%, respectively. In the MRS session, three users gave up when using PC (two on tetrahedron, one on otadecahedron), five gave up when using TB (three on tetrahedron, two on hexahedron), but users completed all sessions when using VR. Table 3 shows the detailed instances of the task failures. Correlation analyses showed no significant relationship between users' familiarity in 3D modelling and their task completion time or workload.

Table 3: Details of the failed instances.

PID	Gender	3D modelling familiarity	Device	Shape
1	Male	Slightly	PC	Tetrahedron
3	Female	Slightly	TB	Tetrahedron
4	Male	Moderately	PC	Tetrahedron
4	Male	Moderately	PC	Octadecahedron
4	Male	Moderately	TB	Tetrahedron
7	Male	Moderately	TB	Tetrahedron
7	Male	Moderately	TB	Hexahedron
11	Female	Moderately	TB	Hexahedron

4.3 Summary

At the end of the experiment, participants were asked to rank the three devices in different tasks based on their subjective preferences. Figure 8 provides a summary of the results of participants' task performance, workload, subjective preference, and success rate.

		PC	TB	VR
Selection	Accuracy	3	1	1
	Workload	2	1	2
	Preference	3	1	2
Manipulation	Time	1	3	1
	Workload	1	1	1
	Preference	1	3	2
	Success rate	2	3	1

Figure 8: Summary of users' task performance, workload, preference, and success rate using PC, TB, and VR in selection and manipulation tasks. Colours in green, yellow and red indicate first, second, and third ranks.

5 DISCUSSION

Our study examined users' task performance and workload on selection and manipulation using three devices that are likely to be used in cross-reality systems: PC, tablet, and VR. We also reported additional findings on users' success rates in manipulation tasks and their subjective rankings of devices in different tasks. In the following sections, we discuss the findings of devices features based on our data analysis results and subjective feedback from participants.

5.1 PC: Great performance in manipulation tasks, but the worst performance in selection tasks

Among three devices, PC was the most frequently used and most familiar device for all participants. As such, participants showed an overall low workload when using PC for all tasks. However, its performance varied in different tasks.

In manipulation tasks, PC showed the best performance and the least workload. During the open discussion, participants also commented that they were familiar with PC controls, and using the mouse cursor to control 3D objects was found much easier than using fingers with tablets: "*I am aware of the object (transform) that I am controlling with the PC cursor, but the occlusion caused by fingers makes it difficult to know where I am clicking when using the tablet*" (P4). Compared to VR, users only need a small range of hand movement to control the target objects when using PC, thus caused relatively less workload, although the difference was insignificant. For example, a participant reported that "*when using VR, I have to make large movements like I am manipulating a real object. With PC and tablets, I can sit down and just move my hand*" (P10).

On the other hand, the selection accuracy of PC was significantly lower than that of tablet and VR. Tablets and VR adopt a direct selection method: users hit the moles through interactions that engage their body movements (fingers or arms). However, the PC selection was indirectly mediated by the mouse. As in line with the findings in previous work [4, 31], indirect methods are often found to be more difficult to learn and operate than direct methods.

5.2 Tablet: Great performance and the least workload in selection tasks, but the worst performance in manipulation tasks

Users' familiarity with tablets seconds that of PCs. However, the results showed that tablet outperformed PC in selection tasks, indicated by the highest task performance, the least workload, and the highest subjective ranking. Users' feedback also indicated that playing WHAC-A-MOLE using tablet was easy and interesting. The direct selection method that engages the use of fingers allowed users to select faster than PC. Meanwhile, finger-based interaction mode also caused less physical workload than VR due to the lower requirements of body movement.

Despite the clear strength, its limitations in manipulation tasks were prominent. It was commented that "*I was not so sure where my finger was tapping, especially when I was making some small adjustments*" (P3). Participants found it particularly difficult to use in manipulation tasks (N=9, 75%). Reasons reported by participants include the limited display area, occlusion caused by the fingers, and the missing physical buttons that led to higher complexity in mode switch compared to PC and VR.

5.3 VR: Balanced performance in all tasks

Although VR was the least familiar device to participants, it showed an overall balanced performance in all tasks and no prominent limitations. Participants mentioned several strengths of VR in 3D object interactions, including the support in spatial awareness, the realism, and the highly alike naturalness in operations compared to interactions in the real world. Participants reported that they can easily move in the virtual space to switch their perspectives, which better helped them observe the virtual objects. They commented that "*being in the same space around 3D objects made the interactions more intuitive and easier than PC and tablets*" (P1). This could have contributed to the fact that no one gave up in the `MRS` session using VR. In addition, participants found the interactions in WHAC-A-MOLE in VR similar to their interactions in the real world and easy to understand. Specifically, they found that the use of the virtual hammer mapped with their mental model. Thus, despite the performance in VR was not the best, many participants ranked the highest on it in selection tasks (N=5, 41.67%).

Compared to PC and TB, participants had a slightly higher workload when using VR, although the difference was statistically insignificant. Participants commented that "*the biggest disadvantage of VR is fatigue*" (P4); and "*VR is interesting but makes me feel tired*" (P12). Some participants (N=4) reported that the weight of the VR HMD has caused some discomfort. Our further analysis showed that users' perceived workload was not associated with their familiarity with the device.

5.4 Selection task difficulty: Significant effect on accuracy and workload in PC and tablet, but not VR

The time-limited selection tasks in WHAC-A-MOLE showed that users' select accuracy decreased and workload increased as the task difficulty increased. For PC and TB, task performance was significantly lower in `L3`, indicating that as the selection task difficulty increases to a critical value, the selection accuracy rate begins to decrease significantly. However, the differences among three levels in VR were insignificant, showing that tasks that are considered difficult in reality (using PC and TB) may not be as hard in VR. Users are capable of completing more complex 3D object selection tasks in VR. In addition, although our experiment results did not show a significant difference between VR and TB in accuracy, the variance in users' task performance was smaller in VR, and there was a tendency for VR to show higher accuracy than TB as the difficulty increased.

5.5 Transform manipulations: PC and VR outperformed tablet; VR showed the highest success rate

For single manipulation tasks (`M`, `R`, `S`), users spent more time using TB in movement tasks and using VR in rotation tasks. The device differences in other tasks were insignificant and all participants completed the single manipulation tasks (`M`, `R`, `S`) within 1-2 minutes. For the longer time in rotation tasks using VR, participants mentioned that the rotation axes were less sensitive than the PC and TB conditions. While our design allowed precise control of the rotation transform, it limited the efficiency. The trade-off is a design factor that need to be considered in future work.

When the manipulation task became complex (i.e. `MRS`), the task completion time increased significantly, and task failures were observed for PC (N=3) and tablets (N=5). We further looked into participant demographic information. The eight task failures occurred with 5 participants (2 females, 3 males), who reported slight to moderate familiarity with 3D modelling. There was no apparent correlation between task failures and users' previous experiences. Although octadecahedrons are more complex than hexahedrons and tetrahedrons in terms of the number of vertices and edges, participants found that it provided more reference points and surfaces, which led to slightly better task performances and lower failure rates.

6 LIMITATIONS AND FUTURE WORK

Our study has some limitations that need to be addressed. First, our results were obtained based on single-user tasks focusing on interaction with 3D objects in reality and virtuality. It does not involve transitions between reality and virtuality or collaboration between users. Effects of the degree of reality/virtuality, information exchange, cooperation mode, and other factors in CR systems should be further explored. In addition, our work mainly contributes to the understanding of the two ends of the reality-virtuality continuum, but not the middle range, such as augmented reality or augmented virtuality. These are also important parts of CR systems and should be explored in future work. Second, we adopted a consistent manipulation method for PC, tablets, and VR based on the three axes for a valid comparison. Direct control that is commonly used in VR was not included in this study, but is of our interest to further investigate in the future. Third, we set a relatively strict target offset in manipulation tasks, as we are motivated by the need for precise control in future cross-reality interaction and collaboration. It is expected that users' task performance may differ if the offset is made more lenient. Future work can explore appropriate offsets for different application areas. Fourth, our results were drawn from a small sample size and all participants were familiar with PC, tablet, and VR. The familiarity has benefited a fair comparison among devices and reduced the influence of experience on the results, but it also led to the limited coverage of participant demographics. Thus, our findings should be generalised with caution.

7 CONCLUSION

In this study, we present a comparative study that investigated the difference between devices (PC, tablet, and VR) on 3D object selection and manipulation tasks. Our research was motivated by the need for cross-device interaction and collaboration in future cross-reality systems. Through our empirical evaluation, we observed significantly different task performances of users using different devices. PC showed significant disadvantages in selection accuracy but worked better than tablets in manipulation tasks. Tablet excelled in selection tasks, but as the task difficulty increased, users' selection accuracy significantly dropped. In addition, users found it challenging to manipulate 3D objects using tablets, especially in movement operations. VR showed a balanced performance in both selection and manipulation tasks. Specifically, we found that unlike PC and tablets, users' task performance in VR was not affected by task difficulty. This indicates that users might be capable of more complex selection tasks in VR than in reality using PC and tablets. Similarly, the success rate of manipulation tasks was the highest in VR when completing complex tasks that involve operations in movement, rotation, and scaling. Our findings contribute to the future design of cross-reality systems that involve selection and manipulation tasks.

ACKNOWLEDGEMENTS

This work is partially supported by the National Natural Science Foundation of China (62207022), Natural Science Foundation of the Jiangsu Higher Education Institutions of China (22KJB520038), and the Xi'an Jiaotong-Liverpool University (RDF-20-02-47, REF-21-01-002). Also, this work is partially supported by the AI University Research Centre and Jiangsu (Provincial) Data Science and Cognitive Computational Engineering Research Centre at XJTLU.

References

[1] L. Chan and K. Minamizawa. Frontface: Facilitating communication between hmd users and outsiders using front-facing-screen hmds. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '17. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3098279.3098548

[2] M. De Marsico, E. Panizzi, F. R. Mattei, A. Musolino, M. Prandini, M. Riso, and D. Sforza. Virtual bowling: Launch as you all were there! In *Proceedings of the International Conference on Advanced Visual Interfaces*, AVI '20. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3399715.3399848

[3] F. Fahmi, K. Tanjung, F. Nainggolan, B. Siregar, N. Mubarakah, and M. Zarlis. Comparison study of user experience between virtual reality controllers, leap motion controllers, and senso glove for anatomy learning systems in a virtual reality environment. *IOP Conference Series: Materials Science and Engineering*, 851(1):012024, may 2020. doi: 10.1088/1757-899X/851/1/012024

[4] L. Findlater, J. E. Froehlich, K. Fattal, J. O. Wobbrock, and T. Dastyar. Age-related differences in performance with touchscreens compared to traditional mouse input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, p. 343–346. Association for Computing Machinery, New York, NY, USA, 2013. doi: 10.1145/2470654.2470703

[5] P. Gebhardt, X. Yu, A. Köhn, and M. Sedlmair. Molecusense: Using force-feedback gloves for creating and interacting with ball-and-stick molecules in vr. In *Proceedings of the 15th International Symposium on Visual Information Communication and Interaction*, VINCI '22. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3554944.3554956

[6] S. Greuter and D. J. Roberts. Controlling viewpoint from markerless head tracking in an immersive ball game using a commodity depth based camera. In *Proceedings of the 2011 IEEE/ACM 15th International Symposium on Distributed Simulation and Real Time Applications*, DS-RT '11, p. 64–71. IEEE Computer Society, USA, 2011. doi: 10.1109/DS-RT.2011.13

[7] J. Gugenheimer, E. Stemasov, J. Frommel, and E. Rukzio. Sharevr: Enabling co-located experiences for virtual reality between hmd and non-hmd users. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, p. 4021–4033. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3025453.3025683

[8] J. Gugenheimer, E. Stemasov, H. Sareen, and E. Rukzio. Facedisplay: Enabling multi-user interaction for mobile virtual reality. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '17, p. 369–372. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3027063.3052962

[9] S. G. Hart. Nasa-task load index (nasa-tlx); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(9):904–908, 2006. doi: 10.1177/154193120605000909

[10] J. Hoffard, X. Zhang, E. Wu, T. Nakamura, and H. Koike. Skisim: A comprehensive study on full body motion capture and real-time feedback in vr ski training. In *Augmented Humans 2022*, AHs 2022, p. 131–141. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3519391.3519400

[11] Y. Im, T. Kim, and E. S. Jung. Investigation of icon design and touchable area for effective smart phone controls. *Hum. Factor. Ergon. Manuf.*, 25(2):251–267, mar 2015. doi: 10.1002/hfm.20593

[12] A. Ishii, M. Tsuruta, I. Suzuki, S. Nakamae, J. Suzuki, and Y. Ochiai. Let your world open: Cave-based visualization methods of public virtual reality towards a shareable vr experience. In *Proceedings of the 10th Augmented Human International Conference 2019*, AH2019. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3311823.3311860

[13] M. Kobayashi, A. Hiyama, T. Miura, C. Asakawa, M. Hirose, and T. Ifukube. Elderly user evaluation of mobile touchscreen interactions. In *Proceedings of the 13th IFIP TC 13 International Conference on Human-Computer Interaction - Volume Part I*, INTERACT'11, pp. 83–99. Springer-Verlag, Berlin, Heidelberg, 2011.

[14] J. J. LaViola Jr, E. Kruijff, R. P. McMahan, D. Bowman, and I. P. Poupyrev. *3D user interfaces: theory and practice*. Addison-Wesley Professional, 2017.

[15] Y. Li, E. Ch'ng, S. Cai, and S. See. Multiuser Interaction with Hybrid VR and AR for Cultural Heritage Objects. In *Digital Heritage 2018*. IEEE, San Francisco, USA, 2018. doi: 10.1109/DigitalHeritage.2018.8810126

[16] Y. Li, E. Ch'ng, S. Cobb, and S. See. Presence and Communication in Hybrid Virtual and Augmented Reality Environments. *Presence: Teleoperators and Virtual Environments*, 28:29–52, 2021. doi: 10.1162/PRES_a_00340

[17] F. L. Luro and V. Sundstedt. A comparative study of eye tracking and hand controller for aiming tasks in virtual reality. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research amp; Applications*, ETRA '19. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3317956.3318153

[18] P. Milgram and F. Kishino. A taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems*, 77:1321–1329, 1994.

[19] M. Nebeling, S. Rajaram, L. Wu, Y. Cheng, and J. Herskovitz. Xrstudio: A virtual production and live streaming system for immersive instructional experiences. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3411764.3445323

[20] O. Oyekoya, R. Stone, W. Steptoe, L. Alkurdi, S. Klare, A. Peer, T. Weyrich, B. Cohen, F. Tecchia, and A. Steed. Supporting interoperability and presence awareness in collaborative mixed reality environments. In *Proceedings of the 19th ACM Symposium on Virtual Reality Software and Technology*, VRST '13, p. 165–174. Association for Computing Machinery, New York, NY, USA, 2013. doi: 10.1145/2503713.2503732

[21] Y. S. Park and S. H. Han. Touch key design for one-handed thumb interaction with a mobile phone: Effects of touch key size and touch key location. *International Journal of Industrial Ergonomics*, 40(1):68–76, 2010. doi: 10.1016/j.ergon.2009.08.002

[22] T. Piumsomboon, G. A. Lee, J. D. Hart, B. Ens, R. W. Lindeman, B. H. Thomas, and M. Billinghurst. Mini-me: An adaptive avatar for mixed reality remote collaboration. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, p. 1–13. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3173574.3173620

[23] I. Radu, T. Joy, and B. Schneider. Virtual makerspaces: Merging ar/vr/mr to enable remote collaborations in physical maker activities. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI EA '21. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3411763.3451561

[24] R. Schlünsen, O. Ariza, and F. Steinicke. A vr study on freehand vs. widgets for 3d manipulation tasks. In *Proceedings of Mensch Und Computer 2019*, MuC'19, p. 223–233. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3340764.3340791

[25] R. A. Settersten. It takes two to tango: The (un)easy dance between life-course sociology and life-span psychology. *Advances in Life Course Research*, 14(1):74–81, 2009. Linked Lives and Self-Regulation. Lifespan - Life Course: Is it really the same? doi: 10.1016/j.alcr.2009.05.002

[26] A. L. Simeone, M. Khamis, A. Esteves, F. Daiber, M. Kljun, K. Čopič Pucihar, P. Isokoski, and J. Gugenheimer. International workshop on cross-reality (xr) interaction. In *Companion Proceedings of the 2020 Conference on Interactive Surfaces and Spaces*, ISS '20, p. 111–114. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3380867.3424551

[27] M. Sra, D. Jain, A. P. Caetano, A. Calvo, E. Hilton, and C. Schmandt. Resolving spatial variation and allowing spectator participation in multiplayer vr. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, UIST '16 Adjunct, p. 221–222. Association for Computing Machinery, New York, NY, USA, 2016. doi: 10.1145/2984751.2984779

[28] R. Strak, K. Yu, F. Pankratz, M. Lazarovici, B. Sandmeyer, J. Reichling, S. Weidert, C. Kraetsch, B. Roegele, N. Navab, U. Eck, and D. Roth.

Comparison between video-mediated and asymmetric 3d teleconsultation during a preclinical scenario. In *Mensch Und Computer 2021*, MuC '21, p. 227–235. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3473856.3473883
[29] P. Tennent, S. Martindale, S. Benford, D. Darzentas, P. Brundell, and M. Collishaw. Thresholds: Embedding virtual reality in the museum. *J. Comput. Cult. Herit.*, 13(2), may 2020. doi: 10.1145/3369394
[30] M. Tiator, C. Geiger, B. Dewitz, B. Fischer, L. Gerhardt, D. Nowottnik, and H. Preu. Venga! climbing in mixed reality. In *Proceedings of the First Superhuman Sports Design Challenge: First International Symposium on Amplifying Capabilities and Competing in Mixed Realities*, SHS '18. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3210299.3210308
[31] N. Vigouroux, P. Rumeau, F. Vella, and B. Vellas. Studying point-select-drag interaction techniques for older people with cognitive impairment. In *Proceedings of the 5th International Conference on Universal Access in Human-Computer Interaction. Addressing Diversity. Part I: Held as Part of HCI International 2009*, UAHCI '09, pp. 422–428. Springer-Verlag, Berlin, Heidelberg, 2009. doi: 10.1007/978-3-642-02707-9_48
[32] J. R. Williamson, M. McGill, and K. Outram. Planevr: Social acceptability of virtual reality for aeroplane passengers. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, p. 1–14. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3290605.3300310