The Effect of Haptics and Spatial Audio on Co-presence in Mixed Reality

A. Agarwal, N. Chavda, T. Chowdhury, G. Desai, C. Ezeji, C. Nookala, M. Zhu, Z. Zhu.

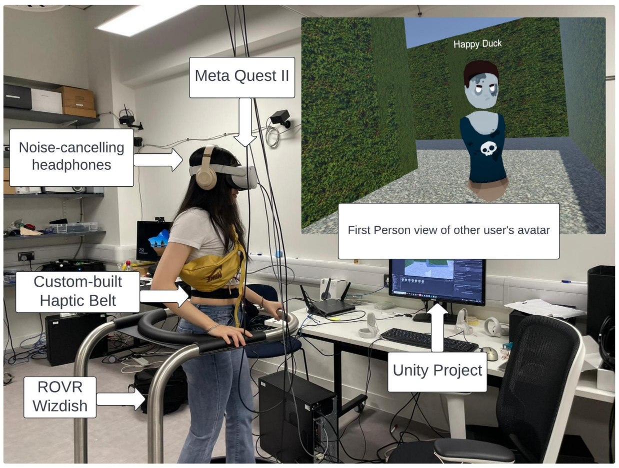

Fig. 1. Live experiment with one user

Fig. 2. Live experiment setup with both users

***Abstract*—Haptic feedback and spatial audio are common senses used within a collaborative Mixed Reality environment to improve presence. This paper aims to compare haptic feedback and spatial audio in collaborative mixed reality, and their effectiveness at improving co-presence, rather than presence. Haptic feedback and spatial audio are implemented as positional cues of other players in a gaming environment. The participants, working in dyads, were tasked with navigating through a maze, using either haptic feedback or spatial audio, to seek out their partner in the maze. The results show that both spatial audio and haptic feedback performs similar in the task. But overall, the higher the co-presence the lower the task completion times.**

***Index Terms*—Mixed Reality (MR), Virtual Reality (VR), Maze, Co-presence, haptic feedback, spatial audio, senses.**

1. Introduction

A Mixed Reality (MR) environment is widely known as one in which both real-world and virtual objects are presented in a single display. This can be achieved with Augmented Reality (AR) and virtual reality (VR) [1]. However, Skarbez et al [2] suggest that, in addition to this, a user must perceive real and virtual objects at the same time, including across different senses. He specifies senses such as olfactory, haptics, audio and taste cues are what truly make a display MR.

Collaboration in MR systems has improved over the past few years, with improvements in communication cues. Presence is defined as the psychological sense of "being there" which is offered with realism [3]. Co-presence is defined as the psychological sense of "being together", at the same time, and therefore, is an important factor for effective communication. Especially in remote collaboration set-ups where the other user is not present in the same place; or in synthetic interfaces where the collaborators are not visible as their true selves. As a result, there is a huge focus on visual cues as a method for improving co-presence. For example, 3D s are more effective than 2D s, utilizing real-time facial expressions and body language alongside their 3D avatars serving as notable examples [4], [5]. On the other hand, senses are not commonly investigated as a factor of co-presence. Although, senses can impact presence. This paper will focus on haptics and audio as the most prevalent senses in the industry currently, and compare the impact they have on co-presence.

This paper was produced with the help of Anthony Steed.

A. Goals and Aims

The experiment investigated co-presence in a gaming environment where visual cues of players are limited. The players used the senses, specifically haptic feedback and spatial audio, to inform their spatial awareness and detect where their partner is located within a collaborative maze environment. Haptic feedback and audio are most frequently used in modern games. The main aim of the game was to seek out their partner with the help of these senses. This paper aims to explore the effect the chosen senses have on the co-presence between the dyads, examining whether one sense has a greater or lesser effect. This was achieved by examining the time taken for dyads to find each other in the maze and the distance traversed by players.

2. Related Work

Wang [4] investigated the use of visual cues in a synthetic interface such as 3D avatars to improve co-presence. They identified other important communication cues such as facial expressions and body language in real-time to replicate the behaviour of humans, giving the 3D avatars a sense of realism. Similarly, Billinghurst and Kato [6] created a device named WearCom which supports 3D spatialized Internet telephony. It utilised spatial audio to give the users who are connected remotely, a sense of feeling as if they are within the same room, and therefore, improved co-presence. This was achieved by determining the distance and direction of the person speaking relative to the listener, and adjusting the volume accordingly to mimic how sound behaves in reality. Spatial audio, a common communication cue in VR environments, provides users with information about the direction, distance, and movement of other participants. Broderick et al [7], emphasised the importance of spatial audio in games. It is a tool used by players to aid their navigational ability and improve decision-making. These examples show how incorporating real-world physics into communication between users can provide a more immersive and accepting MR environment to participants.

On the other hand, Wang et al [8] and Gunther et al [9] showed how haptic feedback can be utilised as a communication cue between participants. Haptic feedback involves the use of tactile sensations, to simulate users' touch senses. Although not seen as a realistic communication cue, it improved performance on tasks working together compared with visual cues alone. However, both experiments concluded that visual feedback was preferred over haptic feedback by participants. Wang et al also found that spatial audio was the less preferred than haptic feedback, but task completion time was the lowest, making spatial audio overall the most effective.

Weisburg et al [10], implemented a compass using vibrotactile sensations in the direction of the true North. These vibrotactile compasses were placed on their arm, and vibrated continuously, to indicate the direction of North. These directional cues were proven to improve the ability to localise and point to internal and external landmarks, even for blind participants. These results demonstrated how vibrotactile compasses can guide users through indoor spaces as an effective directional and awareness cue. This can contribute to users' spatial presence within a MR system, especially where visual cues are limited. The system built for the experiment will implement a custom-made haptic belt similar to Weisburg et al [10], but will provide positional cues of the other player in the maze. Spatial audio will mainly be used as another positional cue than to offer realism. Finally, although visual cues are not needed for the experiment, 3D virtual avatars help players identify when they have found a player in the maze. The outdoor VR maze includes features to provide an immersive experience to users such as sounds of wind.

3. Research Hypotheses

> ***Hypothesis 1.*** *"Spatial audio will be a more effective sensory cue aid in achieving quicker completion times and traversal distances through the maze."*

Building upon the work by Gunther et al [9], spatial audio is the most effective communication cue.

> ***Hypothesis 2.*** *"Spatial audio will be a more usable and natural form of aiding in the positional understanding of others in the same space."*

Spatial audio by the nature of the design, gives a sense of direction in a more continuous frame compared to haptic feedback giving discrete positional cues. Audio allows for subtle differences to be better detected. Also, it provides a better sense of realism [11], [12]. A higher sense of realism consequently improves spatial presence for players. The experiment will also aim to answer the questions outlined by Wienrich et al [13] - does 'spatial presence' refer to the sense of being somewhere, or being with something or someone?

> ***Hypothesis 3.*** *"There will be a negative correlation between co-presence and task completion times"*

This hypothesis asserts that there is a direct link between co-presence and effectiveness in completing the task, also suggesting that co-presence enhances one's ability to complete the task as witnessed in previous collaborative MR studies [14].

4. System Design

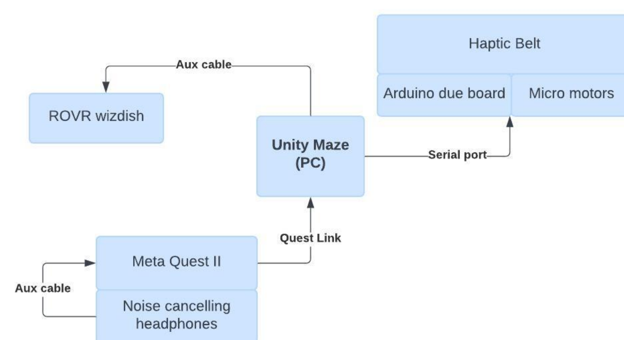

Fig. 3. System design overview

The project required the integration of various hardware and software components to enable the team to collect data accurately while providing a realistic VR experience. These elements were namely a Unity project (in C#), a Meta Quest II headset and noise cancelling headphones, a ROVR Wizdish and a custom-made haptic belt, for each user, as show in Figure 3.

A. Unity Project - 3D Virtual Maze

Users navigated through a Collaborative Virtual Environment (CVE) containing a static virtual maze created in Unity. The design of the maze striked a balance in complexity to maximise traversal possibilities whilst ensuring that users would be able to find each other within a reasonable time frame. The original design of the maze followed a **20 x 20** square grid format.

Within the maze, the team integrated several features to enhance realism and co-presence. These included a 'floating' avatar representing each user with three-point tracking (head, torso and hand tracking), photo-realistic textures, directional light and shadows and animated 3D birds overhead. The avatars in particular were included to enhance the users' sense of presence and allow them to virtually 'discover' one another in the maze. These features were complemented by the implementation of spatial audio which is expanded on below.

The majority of the project was implemented via scripts in C# (aside from some elements of the data logging system which were implemented in Python) with Unity's 2022.3.17f1 Editor and maintained via GitHub for version control.

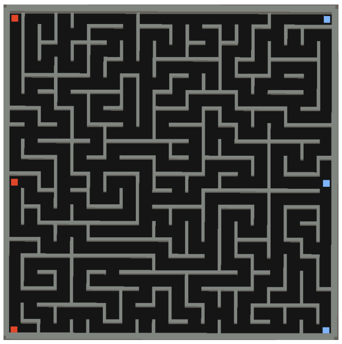

Fig. 4. Maze design with spawn locations

Figure 4 illustrates the maze design used throughout the experiment. The **red and blue** points on the sides and corners of the maze indicate **player spawn points** for *player 1 (red)* and *player 2 (blue)*. For each run of the experiment the the spawn positions are randomized to mitigate any adverse effects the structure of some parts of the maze have on player performance.

Using the unity default scale (1 unit = 1 metre), the dimensions of the maze can be defined. The width of the outer walls was **2m**, the length was **100m** and the height was **10m**. These four bounding walls were the largest in the maze. Due to the overlap of walls to bound the maze, the actual open space was reduced by **1m** on each side, leaving an open maze space of **99 x 99m**. The inner walls were much smaller in size. Each wall was of width **1m** and of height **5m**, the lengths differ to make up the maze structure. Due to the positioning of the inner walls, the path was uniformly narrow around the maze reaching a minimum width of **4m**. Due to this difference in size, at many positions of the maze, the tops of the outer bounding walls could be seen above the maze walls. This gives users minor external cues of their relative position in the maze, but not of the other player, possibly aiding them in their traversal.

B. Ubiq Network Scenes/ Rooms

The system integrated the *Ubiq VR social package* [15] which is an open-source social VR framework built and maintained by UCL's Virtual Environments and Computer Graphics group [16]. Ubiq is implemented as a Unity networking library and was a key component of the implementation as it allowed for multiple users (for the purpose of our experiment, only two) to collaborate in the same virtual space via Ubiq's networking capabilities. This made it crucial to the exploration of co-presence.

When the experiment began, both players were loaded in a *Ubiq network scene* which was hosted on a UCL network server. Both players' applications used Ubiq's messaging service to send positional data such as their avatar's torso, head position and rotation to the other player via the server. The receiver then recreated the avatar as a *remote avatar* object in their own world to make it appear as if they existed in one shared world when in reality both players existed in their own worlds and only had a hologram-like object that streamed the positional data over the network.

The implementation of Ubiq required both users to be in the same *Unity Scene* for the experiment to work properly. Collisions with the maze walls occurred locally, so if even one wall was missing from one player's maze, the other player could see them walk through that same wall on their screen. To enforce this, the experiment ran an identical scene on both players' devices, affirming symmetry across the experiment.

Analysis was done with a "top-down" view in mind and referred to in terms of $\mathbf{x}, \mathbf{y}$ coordinates, despite the y-axis being equivalent to the z-axis within the Unity scene.

C. Logging module

A key component of the Unity maze was the logging capabilities to track the performance of dyads as they traversed the maze. The purpose of the module was to ensure all positional and temporal information could be gauged for each experiment (with analysis done using transformation tools and scripts).

The module was built into the Unity Maze Project as defined in section 4-A and ran in the background during runtime. The module worked in tandem with Ubiq to recognise when two users are connected to the same scene, which acted as a trigger to start logging data. The data logged from the module saved all runtime data in **text files** locally with fields separated by lines and commas.

D. Experimental Space

Dyads carried out the task in the same VR lab, as shown in Figure 2. However, they were spaced out as far apart as possible and wore noise-cancelling headphones to prevent participants from hearing each other, which can detract from the immersive experience.

5. Apparatus

A. Meta Quest II

Alongside the Unity project, the team made use of Meta Quest II [17] headsets to allow the user to experience the maze in 3D through the Quest's high-resolution display, as well as provide a sense of direction through the built-in head-tracking and immersive audio. The latter made the headset integral to providing spatial audio cues to the participants during the spatial audio iteration of the experiment. The Meta Quest was chosen over other VR headset models for several reasons including commercial availability, ease of use and compatibility with Unity and the ROVR.

B. ROVR Wizdish

The ROVR Wizdish was chosen over other conventional forms of player locomotion in VR - for instance, the Meta

Quest controllers - for several reasons. The Wizdish is a virtual reality treadmill designed to translate audio signals to movement velocity in a virtual environment [18]. Specifically, microphones attached to the bottom of the dish provide audio input which is amplified by the sliding motion of the user's feet, while wearing the Wizdish's kick-on overshoes. When combined with the directional input (the user's head rotation) from the Quest II's , this simulates directional movement. Consequently we expected it would help create a more realistic experience, as the action of physically moving one's feet and rotating is akin to natural walking and navigation, allowing for a heightened senses of presence and immersion. Additionally, the mild physical exertion, as one might experience walking naturally, would encourage them to use the sensory cues rather than speeding through the entire maze as they could do easily with conventional controllers. Moreover, users were able to experience the simulation of infinite walking via the Wizdish and their avatar's movement speed through the maze was determined by their physical velocity and augmented via our Unity scripts [19]. Lastly, integrating the Wizdishes with our system helped to overcome the restrictions of running the experiments in a fixed location, expanding our possibilities for the maze design.

C. Headphones for spatial audio

Each VR headset was paired with headphones to deliver real-time auditory feedback, enhancing the spatial awareness of users within the maze. To let the users identify the other user, the system played music within the maze, coming from the other user. Rather than letting the users communicate by talking to each other, playing the music allowed the team to identify and measure accurately the effect of spatial audio on the users' feeling of co-presence.

D. Custom-Made Haptic Belt

The team's experiment featured two custom haptic belts, each integrated with 4 vibration motors controlled by an Arduino Due [20] and a 8-channel relay board [21] for real-time feedback. The motors, shielded by 3D-printed caps, vibrated to user the direction their partner was in and how far they were. The adjustments on the belt allowed the motors to be positioned in the front, right, back and left of the user's waist.

Adapting to resource constraints, the team powered one belt with AAA batteries and the other with an AA battery, using electrical tape for flexible modifications. For safety and comfort, the belts included adjustable velcro straps, ensuring accurate motor positioning for directional feedback. The design mandated wearing the belts over clothing, with vibration intensity carefully monitored for user comfort and safety. The electronics were securely enclosed in a purse, allowing unimpeded user movement.

E. Spatial Audio for Navigation

The players' positions on the maze were recorded as 2D coordinates, while their orientations were represented as bearings from the North direction. Using the other player's position, the Unity maze had a virtual audio source following the avatars, so to each user, it seemed that a song was playing from the other person's avatar. For the user to hear the audio stereophonically, the spatial blend was set to 3D. Additionally, to emulate the effect of walls dampening sound, the program counted the number of walls between the users and dampened the sound by a factor of 15% for each wall. However, users could not differentiate between noises coming from the front or back of the user – unless they turned their heads to figure out the direction of the sound based on the previous and current perceived location of the sound. To make it immediately apparent if the audio is in front or behind the user, the program played a different track depending on whether the other user is in front of or behind the user. To do this, The speaker script calculated the smallest angle between the two vectors (the remote user's avatar direction and the local user's head direction), resulting in an angle between $0°$ and $180°$, where if the angle is $< 90°$, the program determined that the local user is facing the other user, playing track 1. When the angle is $\geq 90°$, the program determined that the local user is facing away from the other user, and played track 2.

F. Haptic Feedback for Navigation

Positional and rotational data of both players were sent to the Arduino from Unity for calculations.

With this data, the Arduino computed the real-time difference in angle and distance for the two users to instruct the correct motors to be turned on and off via the relay board.

For determining the angles between players, consider player 1 and 2's positions as A and B respectively, where $A = (x_1, y_1)$ and $B = (x_2, y_2)$, along with their orientations as α and β respectively. Then the direction each player was facing, $\vec{f_n}$, can be expressed with vectors by substituting their orientations as μ in (1). When the calculated angle between the

$$\vec{f_n} = \begin{bmatrix} \cos(\mu) \\ \sin(\mu - \pi) \end{bmatrix} \tag{1}$$

$$\vec{AB} = \begin{bmatrix} x_2 - x_1 \\ y_2 - y_1 \end{bmatrix} \tag{2}$$

$$\theta = \cos^{-1}\left(\frac{\vec{f_1} \cdot \vec{AB}}{\|\vec{f_1}\| \cdot \|\vec{AB}\|}\right) \tag{3}$$

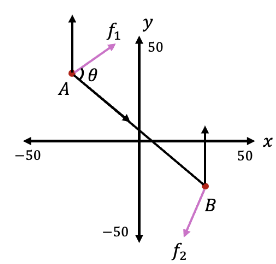

Fig. 5. Calculations for the θ for player 1.

two angles was $\pm 22.5°$ from the directions North, East, South and West then the front, right, back, and left motors vibrate respectively. If the angle was $\pm 22.5°$ from the directions North-East, South-East, South-West and North-West then (two motors) the front-right, back-right, back-left, and front-left vibrated respectively. The same angle calculation is applied to both users.

However, when both users were facing each other, then the North motor vibrated with a frequency. This frequency indicated the distance between the two users - the closer they were the higher the rate of vibrations. This was calculated using a linear mapping from a distance of 0 to a maximum distance of $\sqrt{2000}$ mapped to a delay of 200 to 2000 between each vibration.

6. METHOD

A. Design

The experiment consisted of two conditions. These included:

- **Spatial-audio-only condition:** This experiment employed spatial audio, using music to signal direction and displacement. A unique tune for each user was played, indicating their partner's direction relative to their position in the maze, assisting in navigation.
- **Haptic-only condition:** Participants wore a custom haptic belt equipped with 4 motors. These motors indicated the partner's location relative to the wearer's head orientation

Each condition was run independently, allowing focused analysis of each sensory modality for robust data collection. The maze layout was fixed across pairs for consistent conditions and fair comparison between groups.

Participants had a 10-minute limit to find each other in the maze. If they failed to meet within this time frame or experienced motion sickness prompting a halt, the trial ended, and results were noted as 'Did Not Finish' (DNF) for the participants' safety and accuracy of the conclusions. The sequence of experiments and starting positions in the virtual maze were randomized, as outlined in section 4-A, to reduce performance bias. This ensured participants relied on haptic or audio cues for navigation rather than memory.

B. Participants

24 participants (12 dyads) were recruited from a university student mailing list, all aged 18-25 years. Among them, three dyads were acquaintances, with a gender breakdown of 5 males and 19 females. Approximately 42% reported minimal VR knowledge.

C. Procedure

The study was led by two experimenters. One experimenter monitored each player. Those participants who expressed an interest in the study filled out a pre-experiment questionnaire, read the information sheet, and signed the consent form.

Upon arrival, the players were randomly assigned to a WizDish. Then, they were instructed on how to navigate through the maze using the WizDish, and how to play the maze game by their designated experimenter verbally. Next, a few minutes were given to each user to practice using the WizDish and adjust their HMD. Once the task was clear to the participants, they entered the maze game and the timer was started.

The first experiment, haptic or spatial audio, was randomly selected. Then, the players swapped WizDishes for the next experiment and repeated the game, using the other sense. Each experiment took approximately 20 minutes. For the haptic experiment, the experimenter guided the participant in putting on the haptic belt due to the fragility of the belt, and for the participant's safety. The motors on the belt were adjusted to position them on the front, right, left and back of the participant's torso. Before beginning the experiment, the belt was fastened by the participant and adjusted for their comfort.

Once both experiments ended, the participants were led to desks, where they filled out a post-experiment questionnaire with a PC independently. The participants received a £12 voucher as compensation. Documentation, such as risk assessments, consent forms, advert forms, and information sheets, were prepared in compliance with the UCL CS Research Ethics Committee (reference number - UCL/CSREC/R/16). These documents received multiple reviews by the team's supervisor and experienced researchers to ensure thoroughness and compliance.

D. Measures

To comprehensively assess the impact of haptic and audio feedback on co-presence within a virtual reality environment, the experimental framework collected several measures both before and after the virtual maze task, as well as observations made during the task itself.

1) Pre-Experiment Survey

Prior to participation, a survey captured essential demographic and health-related information from participants. This included occupational background, any health conditions or medications that could influence balance or perceptual capabilities, physical disabilities or limitations, sensory impairments, and a voluntary option to opt out of any part of the experiment. The intent of this survey was twofold: firstly, to ensure participant suitability for the VR task, and secondly, to collect baseline data that might later be correlated with their performance in the virtual environment.

2) During Experiment

Throughout the experiment there were two main sources of data to observe for each run of the experiment with a unique type of data that they provided:

- ***Physical timer*** - Time taken to complete the maze challenge
- ***Logging Module*** - Positional data over the course of traversal

For the physical timer, the team meticulously recorded the time each pair took to complete the maze challenge. Using the *Unity Editor Scene Console*, the position of each user throughout the experiment was tracked and was used to define when users came in contact with each other to stop the timer. This temporal data served as a primary indicator of the efficacy of haptic versus spatial audio cues in fostering a sense of co-presence.

The logging module collected data files for each mode (haptics/spatial). Therefore for each pair of users, there were at least four data files. Data files were marked if they were complete runs of the experiment or not.

Each log file was formatted to have one HEADER line followed by many DATA lines. Both were formatted as follows:

TABLE I
CONTENT IN LOG FILE

HEADER	filename	id1	id2	complete time	comments
DATA	timestamp	x	y	rotation	num walls

Complete time, **x and y** were the most important metrics.

The header was automatically written at the start of each run of the challenge. Once both users joined the same Ubiq room, data lines were written **every second** in both players log files synchronously. When either user left the room to end the experiment, no more log data was written.

Safety was a paramount concern; thus, the team members were vigilant in observing any signs of discomfort among participants and terminated the experiment if either participant could no longer continue.

3) *Post-Experiment Survey*

Upon completion of the task, participants were presented with a survey combining qualitative and quantitative elements. This assessed their experience, specifically focusing on ease of navigation, confidence in their chosen direction, and the effectiveness with which they solved the task. Additionally, the team aimed to evaluate the intuitiveness of the control systems, the participants' perceived sensation of co-presence, and any discomfort or motion sickness experienced during the experiment.

7. RESULTS

Across all testing periods user testing was performed on **12 pairs** of users, each pair yielding at least four data files each (two for each player, two for each mode). Therefore, in total there were a total of **54** data files. Completion times were symmetrical for each pair, therefore there were **27** total completion times. The number of pairs of data for each run of the experiment is shown below:

TABLE II
TOTAL AMOUNT OF DATA COLLECTED

Completion status	Haptics result pairs	Spatial result pairs	**Total**
Success	10	11	**21**
Failure	4	2	**6**
Total	**14**	**13**	**27**

As shown, some runs in the experiment failed to complete, that data was retained but was only applicable for some analysis of the data (e.g. positional traversal data instead of temporal data).

A. *Differences Across Senses*

Taking all the successful completion times, splitting by mode and plotting the box plots, the results are shown in Figure 6.

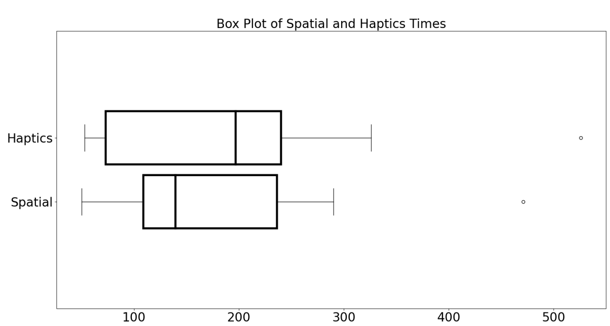

Fig. 6. Box Plots comparing the task completion times

The spatial results show a slight advantage, where both quartiles and median lie below the haptic counterpart, also presenting negatively skewed data whereas the haptics results show positively skewed data. Additionally, the Interquartile Range (IQR) and range are a lot smaller at **127.3** and **240.0** respectively compared to **167.0** and **273.0**. This lower spread and higher consistency in the spatial results potentially indicate this mode has a greater effect on performance and co-presence. In Table III, the mean of each mode (without outliers) indicates that spatial audio yields a slightly lower completion time. Despite this, the standard deviations are high, possibly suggesting limitations in the results by possibly having too few data points to firmly conclude on performance by this metric alone. Performing a **t-test** on the data further affirms this conclusion. The **t-statistic** received was **-0.102**, showing on average the completion time for the spatial audio was slightly higher. However, the **p-value** was **0.920**, which is significantly high. This suggests a high likelihood of observing a difference in the mean if there were no difference between the groups.

TABLE III
TASK COMPLETION TIME COMPARISON

Mode	Mean	s.d.
Haptics	154.7	93.2
Spatial	150.4	74.7

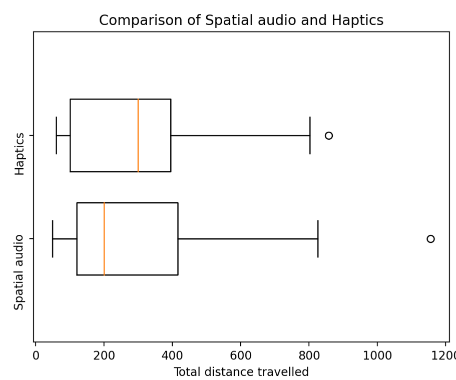

Fig. 7. Box Plots comparing distances travelled by users

Figure 7 compares the distances users travelled in the maze for each condition. A lower median distance travelled in the spatial audio condition compared to the haptics condition implies that on average, participants were able to locate their partners more directly when using auditory cues. However, the near equal range of the IQR for both spatial audio and haptics (295 units rounded to an integer) could suggest that as a navigation aid, both of them have similar learning curves among users. Furthermore, with a **t-test p-value** of **0.865**, there is not enough statistical evidence at a 10% significance level to prove the difference between the distances travelled.

B. *Differences Across Trial Periods*

The majority of test trials went smoothly, yet a notable proportion of participants experienced motion sickness. Given that 42% of participants were new to using HMDs, they showed varying degrees of susceptibility to VR Sickness. The post-experiment survey reveals **7** participants felt sick or nauseous which makes 29% of all dyads. Due to VR sickness, three trial periods had to be cut early and recorded as *DNF*.

C. Post-Experiment Survey Results

TABLE IV
MEAN SCORES FOR EACH MODE FROM THE POST-EXPERIMENT SURVEY.

	Haptics	Spatial Audio
Easiness	3.38	3.42
Confidence	3.38	3.71
Intuitiveness	3.67	3.83
Comfortability	3.67	3.88
Helpfulness	3.67	3.96
Co-presence Rating	3.75	3.63

The results from the post-experiment survey are summarised in Table IV. The mean scores show both these senses had average performance. However, haptic feedback consistently scored lower than spatial audio in all categories. Common problems reported for the haptic experiment included:

- The motors moved out of place, or the belt was loose because sometimes it was difficult to feel the vibrations.
- When the user changed direction, the haptic feedback did not change accordingly due to the lag which confused the user. 2 users explicitly reported experiencing lag or movement delay.

Although not affecting most players, these issues could contribute to lower scores for 'confidence' and 'helpfulness'.

Despite this, 42% preferred haptic feedback over spatial audio overall, 46% preferred spatial audio over haptic feedback, and the remaining found haptic feedback and spatial audio the same. Using the Wilcoxon signed-rank test on the scores for each category showed there was no significant difference between haptic feedback and spatial audio. Overall, the difficulty of the maze was rated average at 3.54. 13% of participants reported frustrations over meeting dead ends and walls in the maze.

8. Discussion

A. Comparisons between spatial audio & haptic feedback

Hypothesis I stems from the findings from previous work by Gunther et al [9], whose findings our experiment builds upon.

Looking at the results of both **t-tests**, there is not enough evidence to reject the null hypothesis. Having **t-test p-values** of **0.920** and **0.865** are both far higher than the **10%** significance level. These values are largely attributed to the low difference in the spread of each set of data as well as the size of the test set being far too small to justify any significant conclusions.

However, the data shown in Figure 6. 7 show that there was a slight advantage in both completion time and distance traversed for the spatial audio mode in the experiment. This suggests that on average, participants were able to locate their partner more directly when using auditory cues. This could be due to the greater dimensional fidelity of the sensory cue itself, with users being able to discern the direction of the other player in a continuous range rather than a discrete one (due to only four motors). The nuance this sensory cue provides gives far more detail on the position of the other user, leading to a higher sense of co-presence.

Additionally, audio as a sensory cue being perceived as a more natural form of navigation as described in section 7-C is in line with the inherent directional nature of human hearing, which allows people to determine the location of sounds in a three-dimensional space accurately. This natural interpretation of the sense could have been less mentally taxing, possibly allowing users to focus more on the details of the sensory cue, rather than interpreting it, thus enhancing the overall sense of co-presence.

Overall, despite **spatial audio** yields slightly higher performance in both task completion times and distance in the experiment, there is not enough evidence to reject the null hypothesis and therefore not enough evidence to accept **Hypothesis I**.Given a future study with more test data yielding similar results, it is likely that would provide more conclusive evidence to accept the hypothesis.

B. Relationship between Presence & Co-presence

Hypothesis II asserts spatial audio gives users a higher sense of spatial presence and is more intuitive. Spatial audio is typically used in virtual environments to increase presence [7], [12]. Overall, there were many positive comments about the maze such as,

"I liked the birds in the maze - added realism"
"Interesting and immersive"
"I experienced a high sense of presence"
"I felt like I was actually in the maze"

This suggests the design of the maze provided an immersive experience and participants experienced high levels of presence overall.

Evidence for **Hypothesis II** is not clear because it can be dependent on the system built. However, Participants described the spatial audio experiment as 'less confusing', 'easier', and 'realistic'. Whereas, participants felt the haptic belt was 'difficult' and 'not clear' to follow. Observations also showed the participants required more time to decipher the vibrations than spatial audio due to more occasional pauses which can distract players from the immersive experience. Therefore, there is some evidence that implies users felt more present with spatial audio than haptic feedback. It was definite that spatial audio was more user-friendly.

Additionally, from users' comments in the post-experiment survey and ratings in the categories in Table IV, one may also postulate that usability had a direct impact on presence and consequently co-presence within the experiment. For instance, 46% and 30% of users indicated that they experienced discomfort or motion sickness during the spatial and haptics sub-experiments respectively, and in six cases this led to aborted or *'DNF'* experiments. Furthermore, many comments on the difficulties experienced by users when using the system as a whole included 'getting lost' a 'slight lag', 'flickering' and the change of direction 'not being clear' with the haptic belt, as well as several comments regarding the lack of stability on the second Wizdish. This is a hindrance to the intuitiveness of spatial audio and users' sense of presence in the experiment overall. And these correlate loosely with some of the lower co-presence ratings given in the response (3 out of 5 and below) suggesting that users who encountered difficulty when using the system experienced a lower sense of co-presence. While this does not provide any clear conclusions regarding **Hypothesis II**, it is important to consider that fair comparisons cannot be made between the haptic belt and spatial audio, and therefore, it is not conclusive that spatial audio makes the user feel more present than haptic feedback.

C. Relationship between Completion Times & Co-presence

A higher sense of co-presence will consequently give users the confidence to make quicker decisions [7]. As well as move more effectively towards their partner in the maze. Better collaboration, and greater co-presence leads to lower task completion times as demonstrated by Grandi et al [14] at least with effective visual cues. This section proves the same for haptic feedback and spatial audio. Linear regression was used to calculate the relationship between the co-presence scores the users gave to the audio and haptic systems and the time they took to complete the game, with the exclusion of DNF data. The analysis revealed a coefficient of -42.08 and a p-value of 0.0961, suggesting sufficient evidence at a 10% significance level to **support Hypothesis III**. This implies that participants with a stronger sense of co-presence were more efficient in locating and reaching their partners.

We chose a 10% significance level due to the variability in participants' experiences, particularly with 29% reporting motion sickness. By excluding data from those who experienced motion sickness, the results showed a coefficient of -58.40 and a p-value of 0.0436, providing stronger support for Hypothesis III. However, this exclusion could introduce bias, as smaller sample sizes may reduce the statistical power. Given the exploratory nature of our study in a relatively new field, a higher significance level was deemed appropriate to identify potential trends for future research.

D. Limitations

1) Unity Integration with ROVR

With the ROVR Wizdish being central to user experience in the experiment, one initial challenge faced when attempting to integrate it with the rest of the project was the fact that a significant portion of ROVR's documentation was either unavailable or implemented via legacy Unity Assets. The team's solution to this was to draw upon a previous Unity implementation of the PacMan game - which also utilized the legacy code, and adapt it to the new implementation. A more robust implementation, one that perhaps integrated better with Unity's more up-to-date assets may be considered for future improvements.

Additionally, while the scripts provided the basic functionality of player movement, there were some observable limitations for instance the latency when translating the player's physical movement on the dish to their movement in the game. This was particularly observable with the second Wizdish, a hardware fault that the team concluded was due to microphone insensitivity, and was observable by some participants. Furthermore, the support rail on the second Wizdish was significantly less stable, impacting the consistency of the user experience across both dishes.

2) Maze

The variation in maze difficulty resulting from random teleportation and deadlocks could have introduced inconsistency and bias into the collected data. In cases where users wasted a significant amount of time trying to overcome these obstacles, they may have experienced increased frustration and disengagement. This could have influenced their decision-making processes, task performance, and feedback. As a result, the reliability and generality of the data may have been compromised.

3) Participant Diversity

One limitation of the experiment stems from the individual differences in sensory perceptions and reactions. For instance, varying levels of hearing and sensitivity to vibrations around the waist among participants could have introduced discrepancies in their responses. Secondly, the occurrence of motion sickness among certain participants might have significantly affected their performance in navigating the maze. Consequently, these phenomena could have influenced the accuracy of our results. Moreover, the majority of participants were recruited from UCL and were female. Consequently, user reactions and responses may vary accordingly.

4) Data Collection

Referring to Table , we only have 21 sets of successful data available for evaluation in our research. This small sample size may introduce sampling bias, potentially result in skewed and less representative conclusions. Moreover, both the task completion time and the co-presence scores may be affected for participants who experienced motion sickness. This confounding bias is unavoidable and can have a stronger influence on the analysis in small samples, potentially leading to inaccurate conclusions.

5) Spatial Audio

Spatial audio presented technical limitations in accurately determining the full 3D location of sound sources. Distinguishing between front and back could have been challenging, potentially confusing users. Additionally, when users were positioned far apart, they may have experienced difficulty in hearing each other, affecting their ability to perceive spatial audio cues effectively.

6) Haptic Belt

The delay in triggering the correct motor vibration due to data transmission could have confused users, particularly if they were turning too quickly. This delay may have required users to spend additional time making decisions. Furthermore, the design of the belt may not have been sufficiently strong and tight as mentioned in the survey feedback, which could have potentially affected its effectiveness in providing directional cues to users.

9. Future Work and Conclusion

While the team was able to effectively integrate the various components of the system, the final implementation was not optimal. For instance, the occasional random rotation of the player's head and torso was observed and tracking of the hands was not available in the game. This was due to the scripts for the Wizdish relying heavily on Unity's built-in physics engine via the motion of Rigid Bodies while the transforms of Ubiq's 'Floating avatar' implementation were accessed directly. For future or similar experiments, these issues could be addressed by restructuring the project to fully utilize the Ubiq package, prompting further modification of the RigidBodyFirstPersonController [22] and replacement of some of the Standard Assets with more up-to-date ones.

Additionally, the team should seek to minimize the bias introduced by the randomization of the players' spawn locations. While the implementation guaranteed that a player never started in the same position twice, thereby eliminating some bias between sub-experiments, it also meant that for specific combinations of spawn locations (the bottom two corners as shown in Figure 4) a significantly easier route could be taken resulting in much lower completion times. While such results were treated as outliers in the analysis, more thought may be given to the maze layout in future projects to make the completion time independent of spawn location, for instance increasing the complexity within certain regions.

As mentioned before, users could not differentiate between sounds from the front and from the back, since there are only two audio sources—one for each ear. In the future, instead of having two tracks to differentiate sounds coming from front and back, the team can have a surround sound system to better achieve spatial audio.

To answer Wienrich et al [13] their question about what spatial presence is defined as in collaborative MR, there were no conclusive results on the relationship between spatial presence and co-presence. This is because there were many issues with the stability of the haptic belt. Several modifications can be made to improve this, including more stable attachment of the motors and wires, perhaps with thicker fabric, as they were observed to occasionally detach during the experiment. Additionally, clearer labelling of the belt to increase ease of use when aligning the motors correctly could be considered.

During the design phase, the team also considered carrying out more iterations of the experiment to explore the impact of sensory cues on co-presence. These included introducing olfactory sensory input alongside or as a replacement for spatial and haptic feedback. Furthermore, more complex multi-sensory input was strongly considered, for instance combining haptics and spatial technology within the same experiment. Having the users complete a different type of task could also provide more data relating to co-presence, such as having two players collaborate or compete with a virtual Jenga puzzle.

Overall, due to an insufficient quantity of test data, there was insufficient evidence to support that lower completion times and traversal distances for seeking their partner in the maze demonstrated higher co-presence. Achieving similar results in a larger-scale experiment could however be used to support the hypothesis. Although evidence for the relationship between presence and co-presence was not comprehensive, there were examples of problems with the haptic belt which hindered presence and thereby may have resulted in lower co-presence scores. Spatial presence within the designed maze game can also be an important factor that impacts the participants' ability to navigate the maze as highlighted by Paul et al [3].

Nevertheless, spatial audio offered the highest levels of co-presence, which aligns and builds upon the findings from previous work.

10. Supplementary Materials

Supplementary materials about our project and the experiment can be found in: https://github.com/Tchowds/Group_Research_Methods_Maze.git

References

[1] P. Milgram and F. Kishino, "A taxonomy of mixed reality visual displays," *IEICE TRANSACTIONS on Information and Systems*, vol. 77, no. 12, pp. 1321–1329, 1994.

[2] R. Skarbez, M. Smith, and M. C. Whitton, "Revisiting milgram and kishino's reality-virtuality continuum," *Frontiers in Virtual Reality*, vol. 2, p. 647997, 2021.

[3] P. V. Schaik, T. Turnbull, A. V. Wersch, and S. Drummond, "Presence within a mixed reality environment," *Cyberpsychology and Behavior*, vol. 7, 2004.

[4] R. Wang and X. Wang, "Mixed reality-mediated collaborative design system: Concept, prototype, and experimentation," in *Cooperative Design, Visualization, and Engineering: 5th International Conference, CDVE 2008 Calvià, Mallorca, Spain, September 21-25, 2008 Proceedings 5*, pp. 117–124, Springer, 2008.

[5] F. Moustafa and A. Steed, "A longitudinal study of small group interaction in social virtual reality," in *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, VRST '18, (New York, NY, USA), Association for Computing Machinery, 2018.

[6] M. Billinghurst and H. Kato, "Collaborative mixed reality," in *Proceedings of the first international symposium on mixed reality*, pp. 261–284, 1999.

[7] J. Broderick, J. Duggan, and S. Redfern, "The importance of spatial audio in modern games and virtual environments," in *2018 IEEE Games, Entertainment, Media Conference (GEM)*, pp. 1–9, IEEE, 2018.

[8] J. Wang, A. Chellali, and C. G. Cao, "Haptic communication in collaborative virtual environments," *Human factors*, vol. 58, no. 3, pp. 496–508, 2016.

[9] S. Günther, S. Kratz, D. Avrahami, and M. Mühlhäuser, "Exploring audio, visual, and tactile cues for synchronous remote assistance," in *Proceedings of the 11th pervasive technologies related to assistive environments conference*, pp. 339–344, 2018.

[10] S. M. Weisberg, D. Badgio, and A. Chatterjee, "Feel the way with a vibrotactile compass: Does a navigational aid aid navigation?," *Journal of experimental psychology: learning, memory, and cognition*, vol. 44, no. 5, p. 667, 2018.

[11] P. Voss, "Auditory spatial perception without vision," *Frontiers in Psychology*, vol. 7, 2016.

[12] G. Corrêa De Almeida, V. Costa de Souza, L. G. Da Silveira Júnior, and M. R. Veronez, "Spatial audio in virtual reality: A systematic review," in *Proceedings of the 25th Symposium on Virtual and Augmented Reality*, SVR '23, (New York, NY, USA), p. 264–268, Association for Computing Machinery, 2024.

[13] C. Wienrich, P. Komma, S. Vogt, and M. E. Latoschik, "Spatial presence in mixed realities–considerations about the concept, measures, design, and experiments," *Frontiers in Virtual Reality*, vol. 2, 2021.

[14] J. G. Grandi, H. G. Debarba, and A. Maciel, "Characterizing asymmetric collaborative interactions in virtual and augmented realities," 2019.

[15] S. J. Friston, B. J. Congdon, D. Swapp, L. Izzouzi, K. Brandstätter, D. Archer, O. Olkkonen, F. J. Thiel, and A. Steed, "Ubiq: A system to build flexible social virtual reality experiences," in *Proceedings of the 27th ACM symposium on virtual reality software and technology*, pp. 1–11, 2021.

[16] Virtual Environments and Computer Graphics Group, UCL . https://ucl-vr.github.io/ubiq/. Accessed 01/2024.

[17] Meta. https://www.meta.com/gb/quest/compare/. Accessed 01/2024.

[18] . ROVR Wizdish Team. https://www.wizdish.com/rovr-specification. Accessed 02/2024.

[19] COMP0013 Group 9. https://github.com/Tchowds/Group_Research_Methods_Maze/tree/chidinma. Accessed 01/2024.

[20] A. O. Store, "Arduino Due." https://store.arduino.cc/products/arduino-due. Accessed 02/2024.

[21] Components101, "5V 8-Channel Relay Module." https://components101.com/switches/5v-eight-channel-relay-module-pinout-features-applications-working-datasheet. Accessed 02/2024.

[22] Unity Manual. https://docs.unity3d.com/560/Documentation/Manual/HOWTO-InstallStandardAssets.html. Accessed 01/2024.