Comparison of techniques for mixed-space collaborative navigation

Aaron Stafford **Bruce H. Thomas** **Wayne Piekarski**

Wearable Computer Lab, School of Computer Science
University of South Australia,
Mawson Lakes Blvd, Mawson Lakes, South Australia, 5095
Email: {aaron.stafford|bruce.thomas|wayne.piekarski}@unisa.edu.au

Abstract

This paper describes the results of two studies conducted to determine the role of visual cues for a collaborative navigation task in a mixed-space environment. Both studies required a user with an exocentric view of a virtual room to navigate a fully immersed user with an egocentric view to an exit. The first study compares natural hand-based gestures, a mouse-based interface and an audio-only technique to determine their relative efficiency on task completion times. The follow-up study compares natural hand-based gestures against a mouse-based interface in a scenario in which participants are unable to communicate verbally.

The results show that visual cue-based collaborative navigation techniques are significantly more efficient than an audio-only technique. The results also show that natural hand gestures are more expressive and lead to quicker completion times in situations where verbal communication is not possible.

Keywords: Mixed-space collaboration, user study, virtual reality, tabletop interaction.

1 Introduction

Mixed-space collaboration typically involves a number of people viewing the same problem space from different perspectives and from different levels of immersion to solve a problem (Grasset et al. 2005). Collaborative navigation is required in situations where navigational targets or conditions of the surrounding environment are constantly changing and therefore can not be predicted or represented by a computer system.

Previous research has demonstrated the effectiveness of collaborative mixed-space navigation to improve navigation times (Grasset et al. 2005). The study required a participant with an exocentric viewpoint of a maze to guide a fully immersed participant in the maze towards the exit using only verbal communication. The number of options at any decision point was at most four: "go straight ahead", "go left", "go right", and "go back the way you came". This research showed that exocentric-egocentric collaboration is significantly more efficient for navigational tasks compared to single person navigation. In more complex real world scenarios, the number of alternatives is arbitrary. As the number of alternatives increases, it should become increasingly difficult and therefore more time consuming to remotely navigate a person using only voice commands.

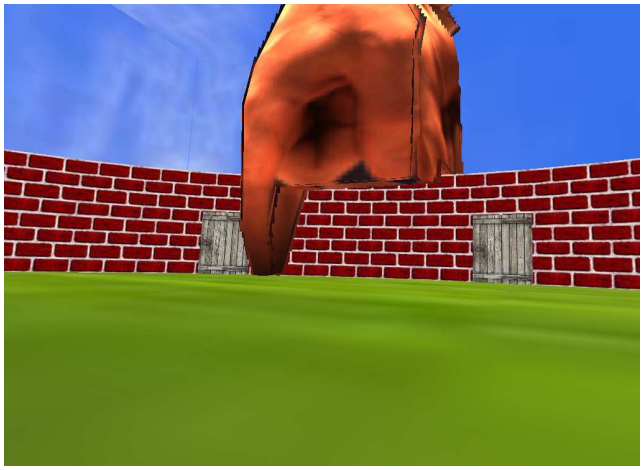

Figure 1: The round virtual room in which the VR users were immersed. Also visible is the reconstructed hand of HOG table participant.

In such examples of collaborative navigation, verbal navigation instructions need to be given with respect to the person with the egocentric view of the navigation space. Therefore, the person with the exocentric viewpoint of the navigation space first determines the spatial relationship between the person with the egocentric view and their goal before describing it to them verbally, for example, "The exit is on your right".

God-like interaction has previously been presented as a metaphor for communication between people located indoors, using a 3D reconstruction tabletop display, and people located outdoors using an outdoor augmented reality (AR) system (Stafford et al. 2006). The 3D reconstruction tabletop display (HOG table), as seen in Figure 2, is capable of capturing users' hand gestures and tangible prop interaction and conveying the 3D information to people outdoors using outdoor AR systems. Using this system, a person with an exocentric view of the navigation space can guide the person with the egocentric view to their goal without building a mental model of the spatial relationship between the person with the egocentric view and their goal.

This paper presents two studies conducted to determine the role of visual cues for mixed-space collaborative navigation. These studies evaluate the following:

1. The effectiveness of using visual cues compared to audio-only cues for navigation purposes.
2. The effectiveness of using god-like interaction compared to mouse-based input for situations

where communication through audio is not possible.

2 Background

There has been a wide range of work in the area of collaborative 3D environments such as AR and virtual reality (VR). Systems such as Studierstube (Schmalstieg & Hesina 2002) have been developed which allow multiple users to work together and edit 3D models in real-time, sharing a distributed scene graph between multiple application instances. With users present in the same room, they can use traditional forms of communication such as speech to coordinate their actions easily. When users are not present in the same location, collaborative tasks are more difficult and other methods of communication need to be explored.

In order to help to better understand interactions in different types of environments, Poupyrev et al. (Poupyrev et al. 1998) created a novel classification for Virtual Environments (VE) manipulation metaphors. The classification separates metaphors into egocentric or exocentric depending on the user's viewpoint. Exocentric are those metaphors in which users have an external or god's eye view looking down onto the world. Egocentric metaphors are typically used in immersive systems and place the user directly in the environment.

Collaborators in a mixed-space environment have been shown to experience better shared understanding using a similar egocentric frame of reference (Schafer & Bowman 2004). This is because the collaborators have a similar understanding about the spatial relationship of objects in the environment. However, mixed-space collaboration that combines an exocentric navigator with an egocentric pilot is more efficient for navigational tasks when navigation is restricted to a 2D plane (Kopper et al. 2006, Grasset et al. 2005, Schafer & Bowman 2004). This is due to the exocentric user having a better understanding about the spatial relationship of objects in the scene. So while participants do not share the same understanding of the spatial relationship of objects they are able to utilise the combined understanding (Brown et al. 2003) to better complete tasks.

A study in a collaborative virtual environment to determine the influence perspective has on collaborative navigation found that in a 3D space, while location time is quicker with an exocentric viewpoint, navigation becomes increasingly difficult the more exocentric the viewpoint (Yang & Olson 2002). Participants piloted virtual submarines to locate flashing targets; One participant was the driver, the other the navigator. The experiment tested four viewpoint on a range from egocentric through to exocentric. The results showed that the more exocentric the viewpoint the faster the search times. But the more exocentric the viewpoint the longer the travel times. These results indicate that while the exocentric viewpoint makes destination identification easy, the mental rotation required to navigate the driver in terms of their perspective incurs significant time costs.

Collaboration at different scales (Leigh & Johnson 1996, Grasset et al. 2005, Billinghurst et al. 2001, Bowman et al. 2004, Nakanishi et al. 2004) enables two or more users to leverage their different viewpoints to complete tasks in a more efficient manner. A common approach to multi-scale collaboration is to provide one user with an egocentric view of the world and another user with an exocentric view of the same world (Leigh & Johnson 1996). The exocentric viewpoint is useful for obtaining a bird's eye view of the mixed-space world. Therefore, the exocentric user is in a good position to provide the egocentric users with instructional information such as navigation information.

Nakanishi et al. (Nakanishi et al. 2004) use the term transcendent communication to refer to the interaction between people with a bird's eye view of an area of interest and those at ground level in the area of interest. A study conducted by the authors found that the bird's eye view was effective for understanding the spatial movements of crowds. The authors also found that users with a bird's eye view were able to effectively assist people at ground level in understanding their surroundings.

Similarly, a Greek god metaphor has been used to describe the relationship between a desktop user and a fully immersed VR user (Holm et al. 2002). In this research, the users work together to build a 3D environment. To the VR user, the desktop user appears as a giant hand interacting with the environment. The VR user is known as the hero and has the ability to lift massive objects that would not be possible in the real world. An interesting benefit of this work is that while both users design the environment, the immersed user also immediately experiences the environment and provides practical feedback about the usability to the desktop user. This system is limited to working with a set of prefabricated objects. The terms deity and mortal have been previously used to described the exocentric and egocentric roles assumed by users in a collaborative virtual environment (Leigh & Johnson 1996).

The use of god-like interaction (Stafford et al. 2006) for providing navigational information is analogous to navigation via landmarks as the reconstructed hand can be used as a landmark to be navigated to. Landmarks in the real world assist building geographical knowledge of surroundings by providing spatial references to identifiable locations (Evans et al. 1984). It is therefore easier to learn a route in an environment with useful landmarks than in an environment void of landmarks (Tlauka & Wilson 1994). Landmarks are just as applicable in large virtual worlds as they are in the real world. An experiment to determine whether or not people use real world wayfinding strategies in large virtual worlds, measured participant performance of a complex search task conducted in a number of different large virtual worlds (Darken & Sibert 1996). The results show that participants became disoriented and have trouble completing the task when the virtual world is void of any typical physical world navigation cues.

Vinson takes the best practices for landmark design in the real world and created 13 guidelines for landmarks to support navigation in virtual environments (Vinson 1999). Vinson's 5th guideline is that virtual landmarks should be visible at all navigable scales. In virtual worlds, users are not bound by the physical limitations imposed on them in the real world and so flying in virtual worlds is an efficient form of travelling. Additionally objects such as landmarks are selectable in virtual worlds, therefore a user can instantaneously travel to a landmark in a virtual world with minimal effort if the landmark is visible from a long distance away. However, it has been shown that instantaneous travelling can result in disorientation (Stoakley et al. 1995).

3 Experiment design

The study took place in the VR Lab at the School of Computer and Information Science. The HOG table communicates over a high bandwidth reliable network (using a 1Gbps LAN) to a computer with a 64 bit 2.4Ghz AMD Athlon processor and 512MB of RAM. An Nvidia GeForce 6800 GT drives an 800x600

Figure 2: HOG table participant using their hand to point to a location in the virtual world.

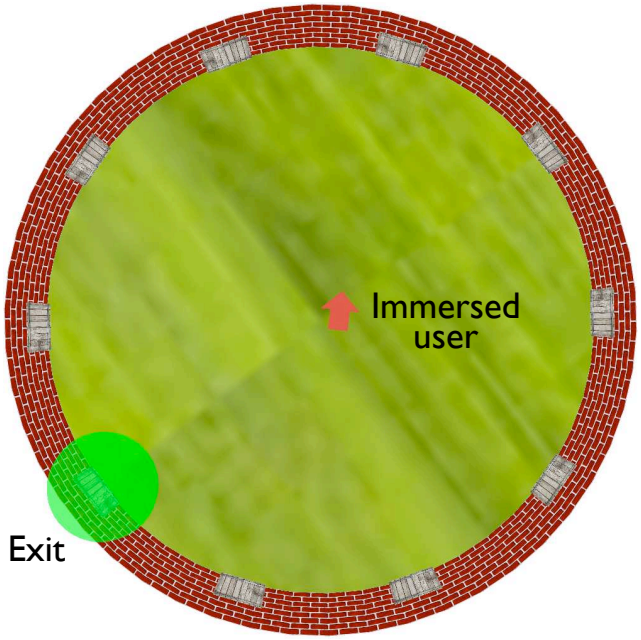

Figure 3: The view of the HOG table participant, the red arrow represents the immersed participant's location and head orientation, the opaque green circle above one of the doors highlights the exit to the room.

I-glasses HMD that has a horizontal FOV of approximately 26 degrees. A Polhemus 3Space Fastrak magnetic tracker tracks the position and orientation of the HMD and two hand controllers with 6 degrees of freedom. The hand controllers each have 5 buttons that write a signal to a serial bus when pressed. A custom application written in C++ and OpenGL facilitates communication with the network, tracking and serial devices. The application also maintains a simple scene graph. The objects captured by the HOG table are rendered as a node in the scene graph.

The task was conducted in pairs. One participant was fully immersed in a virtual room using the VR system described above. The virtual room consists of a round room that appears to be 10m in radius (see Figure 1). Around the wall of the room are a number of doors. All doors look identical and the only way to find the exit without assistance is to test every door. Navigation through the virtual room was achieved via a hand controller using four (forward, back, strafe left, strafe right) of the five available buttons. Pressing a button would cause the viewpoint to be translated in one of the directions relative to the view vector of the immersed participant. It was not possible for the participant to fly up and down. The immersed participant's task was to find the exit door with assistance from the HOG table participant.

HOG table participants saw a top-down view of the room that the immersed participant was in (see Figure 3). A red arrow representing the location and head orientation of the immersed participant was visible to the HOG table participant. The HOG table participant could also see a semi-transparent green circle over the exit door.

The navigation task was performed over three conditions:

Audio-only: The HOG table participant is restricted to only issuing verbal commands to guide the immersed participant to the final destination.

Mouse-based: The HOG table participant uses a mouse to control a cursor. Where they clicked on the 2D top-down view of the room, a small blue dot appeared, while in the immersive view a hand appeared (see Figure 1). The hand stayed at the clicked location until a new location was clicked.

Gesture-based: HOG table participants use hand gestures to guide the immersive participants to the exits. HOG table participants were shown the various ways in which the interface could be used. The following examples were explained to the user: pointing to the exit, pointing left or right, pointing to the side of the exit, pointing to either side of the exit, and dragging a finger to trace out a path that could be followed. The end result for pointing gestures is very similar to the hand in Figure 1, but in the gesture-based case the reconstruction was updated 7 times a second. The rate is limited by image processing time of the HOG table.

HOG table and immersed participants were able to communicate audibly for all of the conditions. The HOG table was on a height adjustable motorised table. At the start of each condition the HOG table participant was asked if they would like the height of the table adjusted.

To determine which participant would be the immersed participant and which would be the HOG table participant, both were asked if either had a preference in case one knew that they were susceptible to motion sickness. If no consensus could be reached the roles were determined by the toss of a coin.

The pairs experienced each of the conditions until they indicated they were comfortable with using the system, i.e., how to navigate, how to provide navigational information, and what to expect to see. For each of the conditions, the HOG table participant's task was to guide the immersed participant to the exit 20 times. The number of doors was randomised nad changed each time the immersed participant found the exit. The number of doors was always between 3 and 12 inclusive, therefore there were 10 different rooms and each room was experienced twice per condition.

For each room, the number of doors and immersed participant's initial heading relative to the exit were recorded by the system. HOG table participants were video-taped during the study to capture the words used during the different conditions and also to capture the various techniques used so that they could be analysed with respect to the results.

At the end of the study participants were asked

to complete a questionnaire, with both participants having different questionnaires based on their role.

HOG table participants were asked to rank the following questions on a 7-point Likert scale:

1. We had problems and struggled to complete the task.
2. My partner always understood my directions.
3. It was easy to get my partners attention.
4. It was easy to correct mistakes made by my partner.
5. I could easily communicate my intentions.
6. I had to concentrate very hard to do the task.

Immersed participants were asked to rank the following questions on a 7-point Likert scale:

1. We had problems and struggled to complete the task.
2. I always understood my partners directions.
3. I always knew which door to go through.
4. It was easy for my partner to grab my attention.
5. I always understood when my partner was trying to correct my mistakes.
6. I could easily understand my partner's instructions.
7. The task was easy to complete.
8. The task required little effort.

After the questionnaires, participants took part in an informal interview regarding their experiences and the techniques they used. Some extracts from the interviews are used throughout this paper to help explain the observed results.

3.1 Hypothesis

The following hypothesis were formed before conducting the study:

1. Visual cue-based navigation would be more efficient than audio-only navigation.
2. Gesture-based navigation will be easier than mouse-based navigation.

3.2 Pilot study

The initial pilot study involved three pairs. This pilot study revealed that flight in one direction was all that was typically used for navigation. Therefore for simplicity, only one button was required for navigation and this was mapped to a flying action along the viewing vector.

In the user study room, a monitor attached to the VR system displayed the same view as the immersed participant was experiencing. HOG table participants would look at this screen to see what the immersed participant could see. Participants would modify their behaviour based on what they knew the immersed participant could see. Therefore it was decided to make this part of the experiment as well. The display was deliberately placed such that the HOG table participant could easily see the display, and it was explained to them that it showed the immersed participant's view of the room.

The immersed participant would wait until they had identified the exit before beginning to fly from their starting position. The participant would rotate around on the starting position while searching for the exit, but they typically did not begin moving forward until they were reasonably sure they knew which door the exit was. Therefore, the system was modified to record the time before immersed participants began to fly, and for the remainder of this paper this recorded time is referred to as "locate time". The time from beginning to fly until the participant found the exit was also recorded, and for the remainder of this paper this recorded time is referred to as the "travel time".

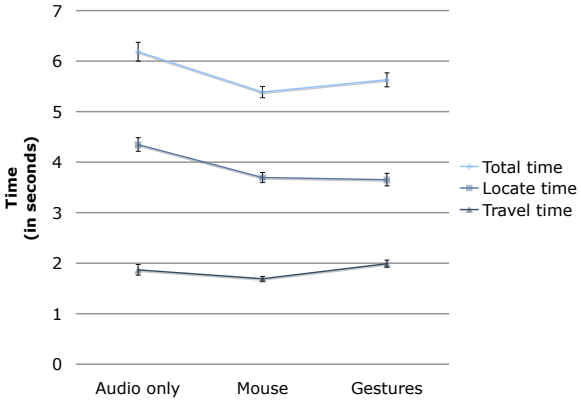

Figure 4: A comparison of the average times to locate and travel to the exits under the various conditions.

4 User study

The main user study was conducted in the same fashion as the pilot except:

- only a single button on the hand controller was used for navigation,
- a monitor was positioned close to the HOG table so HOG table participants could see the immersed participants' view,
- "locate time" and "travel time" were recorded,
- groups experienced the same three conditions in a different order to compensate for learning effects.

4.1 Results

The study involved 12 groups of 2 participants made up of 19 males and 5 females. All of the participants worked with Windows-based computers for more than 15 hours per week. Of the immersed participants, 58% had less than 1 hour previous experience with VR, 33% of the immersed participants had between 1 and 5 hours previous experience with VR, and only 1 immersed participant had more than 5 hours of previous experience with VR. In 3 of the 12 groups the participants did not know each other.

4.1.1 Task completion times

The average locate time using an audio-only condition (4.35 seconds) is longer than with the mouse (3.7 seconds) or gesture condition (3.65 seconds) (see Figure 4). This is as expected since the gesture and mouse conditions both provide a visual confirmation of the exit. Without a visual cue, more dialogue and hence more time is required to locate the exit. A one-way-within subjects ANOVA on the total times with

a significance level of $\alpha = 0.05$ reveals a significant difference with $p < 0.05$. A post hoc t-test analysis (a Bonferroni correction of alpha value to 0.01), was performed for the three conditions, gesture and mouse conditions ($p > 0.01$), audio-only and mouse ($p < 0.01$), and audio-only and gesture ($p < 0.01$). The results show the significant effect the visual cue-based approaches have over an audio-only approach with both the gesture and mouse conditions significantly faster than the audio-only condition. The results also show that there was no significant difference between the efficiency of the gesture and mouse conditions. This supports our hypothesis that visual cue-based navigation is more efficient than audio-only navigation.

For the audio-only condition, HOG table participants would typically guide the immersed participant starting with a direction such as "left" or "right", or quite often with a rough number of degrees such as "turn 90" or "turn 180." As the immersed participant rotated, the HOG table participants would typically offer encouragement such as "keep going". When the immersed participant was about to look at the door or just as they did, the HOG table participants would say "stop" or "that's it." At this point the immersed participant would begin travelling to the exit. HOG table participants would often correct any mistakes at this point by saying something like: "no, it is the next one on your left." This is evident from the participant feedback given during the post task interview, where one participant responded:

> "I found basically the hand gestures and the mouse about the same, they just really helped me find exactly what door it was. With voice it was just sometimes hard to pinpoint when there was quite a few doors."

This made the audio condition the least efficient condition overall. However, the average travel time for the gesture condition (1.99 seconds) is greater than both the audio-only (1.87 seconds) and mouse conditions (1.69 seconds) with very little variation. This is likely due to the nature of the reconstructed gestures. It was common for immersed participants to comment on the quality of the real-time reconstructed gestures, which were also considered less accurate than the mouse condition by both participants.

For the mouse condition, HOG table participants would typically initially employ the same technique as for the audio-only condition, such that they would start the immersed participant looking in a particular direction such as left or right. By this time the HOG table participant would have had enough time to click on the exit. The immersed participant would see the exit that was being pointed to and navigate to it.

For the gesture condition there were two distinctly different approaches taken by HOG table participants. The first is very similar to the mouse condition. HOG table participants provided a direction to start looking, and as the immersed participant started looking around the HOG table participant would reach down and point to the exit. The other approach was for the HOG table participant to initially point in the field of view of the immersed participant and trace a line for the immersed participant to follow to the exit. These two techniques are discussed later with respect to other results.

The standard error for the locate and travel times are also presented in Figure 4. It is clear that there is more variation in the gestures condition (0.14s) compared to the mouse condition (0.11s). The larger variance could be attributed to the different approaches that HOG table participants took as described above. Whereas the HOG table participants for the mouse condition took a more consistent approach.

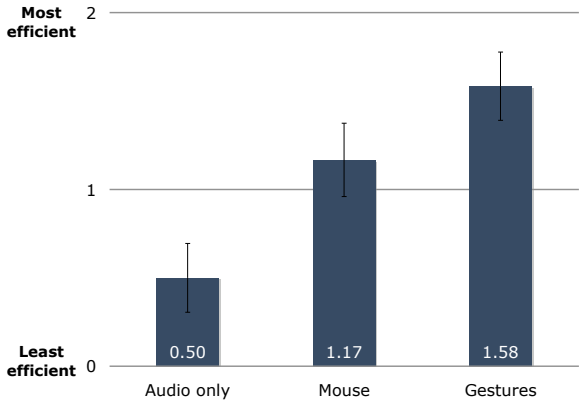

Figure 5: HOG table participants were asked to order the conditions in order from most efficient to least efficient.

In the questionnaire after the experiment, both participants were asked to rank the conditions in order of which they thought was the most efficient to which they thought was the least efficient. A condition that was ranked as the most efficient was given 2, the next most efficient 1 and the slowest 0. In Figure 5 the HOG table participants ranked the gesture condition as the most efficient (1.58), followed by mouse condition (1.17) and then the audio-only condition (0.5). This results seems to be due to the 1:1 relationship between the location of the hand and the point where they wanted the immersed participant to move to. Compared to the mouse which requires some thought about where the mouse currently is and where it has to be moved to. One participant's comments:

> "With the hand it is natural because you just know where it is whereas with the mouse it is another layer of abstraction because it is an abstract interface so, regardless, you have to still move the mouse to the new location you want and when you over shoot it is harder to correct."

Another participants said:

> "I think it is because with pointing you can kind of just know where it is and you can just point whereas with the mouse you've got to look at where it is, making sure the mouse gets there."

The immersed participants had a different view of the efficiency of the conditions. The difference between the mouse and the gesture condition is less obvious than the HOG table participants' responses to the same question (see Figure 6). In the post task interview most immersed participants saw little difference between the mouse and gesture as in both cases the visual cue appeared as a big hand. Perhaps the reason that immersed participants failed to notice the same efficiency as the HOG table participants was due to what immersed participants often saw of the real-time reconstruction. HOG table participants would experiment more with the gesture interface sometimes without realising the consequences, one HOG table participant describes the following example:

> "What I was trying to do was put two fingers over the doorway so there was no confusion about it but then I looked up at one stage

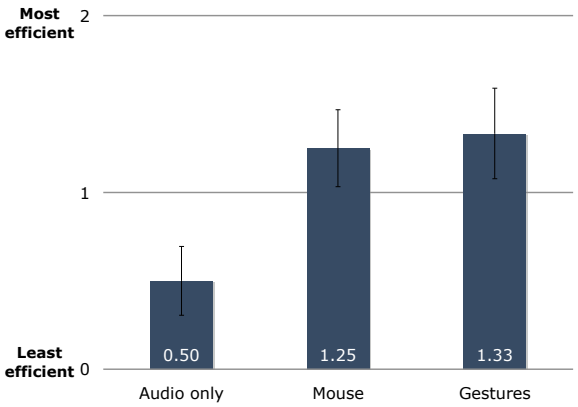

Figure 6: Immersed participants were asked to order the conditions in order from most efficient to least efficient.

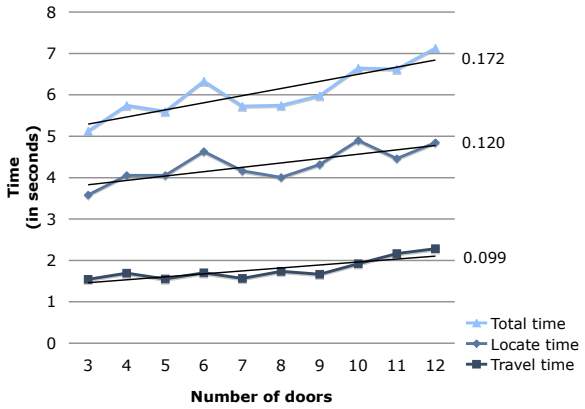

Figure 7: The effect of the number of doors on task times for the audio condition.

> and because of the way the cameras were occluding it was just one big block somewhere and when I looked down he [the immersed participant] was walking into the wall because he was standing in the occluded area. That is why I changed the way I was doing it half way through to putting my finger down and pulling it out as soon as he started heading in the right direction. Which helped I think."

While gestures seemed more efficient to the HOG table participant, they were also experimenting more and this lead to confusion for the immersed participants.

4.1.2 Number of doors

Figure 7 shows the effect the number of doors had on navigation times in the audio-only condition. The graph shows that as the number of doors increases so does the travel time, locate time and therefore the total time. This is expected as for the immersed participant a door that is initially described to be "on the left" becomes ambiguous when there are more doors in the room as on the left there will be two or more doors filling the field of view, and therefore it takes further interaction between the immersed participant and the HOG table participant to resolve the ambiguity.

For the mouse condition the number of doors has very little if any effect on the locate time or the travel time (see Figure 8). The line of best fit through the total time data shows a slope of just 0.005 which for all purposes is essentially flat, particularly when compared to the slope of the total time for the audio-only condition (see Figure 7) which is 0.172 (34 times greater). Therefore we conclude that in the mouse condition the number of doors does not effect task completion times.

As with the travel time for the mouse condition, the travel time for the gesture condition is also effectively completely flat (see Figure 9). Suggesting that, unlike the audio condition, the effect the number of doors has on travel time in the gesture condition is negligible. However the slope of the line of best fit for the locate time in the gesture condition (0.078) is about 9 times greater than the slope of the locate time for the mouse condition and about 0.6 times the gradient of the slope of the locate time for the audio-only condition. This indicates that the number of doors did effect the locate times in the gesture

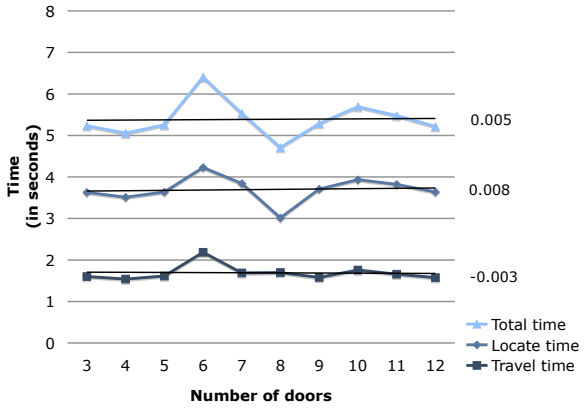

Figure 8: The effect the number of doors on task times for the mouse condition.

condition, but not as much as for the audio condition. This is likely to be attributable to the different techniques used by the HOG table participants as described previously. The technique to direct immersed participants from their starting position to the exit must lead to an increase in time and people "experimenting" with the interface.

In the three graphs comparing the number of doors to the task completion times (see Figures 7, 8 and 9) there seems to be common features. For example, all three graphs seem to spike and door 6 and drop again sharply at door 8. This is a cause for concern as it may indicate that there is something about the set up of the experiment that influenced the results over all conditions.

Figure 10 is a graph that shows the average number of degrees from the start orientation to the exit door for each number of doors. In the room with 6 doors the exit is on average 114 degrees from the start orientation. This is 15 degrees more than the next highest (room with 12 doors) and 32 degrees more than the total average for all rooms. Therefore, the spikes in the previous graphs are attributable to the large angle between the immersed participant's starting orientation and the direction to the exit door. Figure 10 shows a strong dip in the room with 8 doors. The number of degrees between the immersed participant's start orientation and the exit door is roughly equal to the second lowest angle (68 degrees). This means that for rooms with this number of doors the exit door would have been in or nearly in the participants' immediate field of view more often than not,

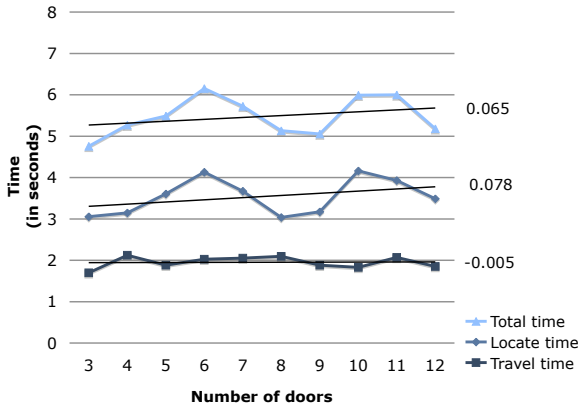

Figure 9: The effect of the number of doors on task times for the gesture condition.

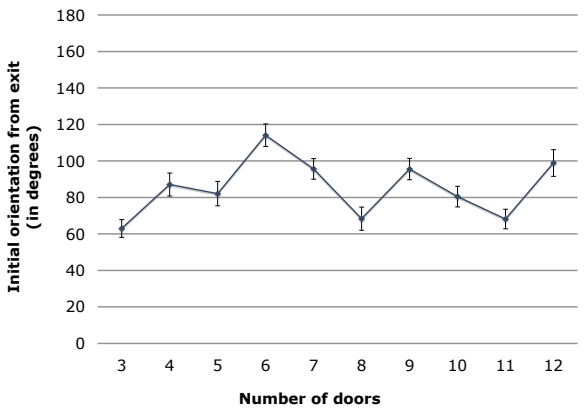

Figure 10: The average number of degrees from the start orientation to the exit door for each number of doors.

therefore leading to lower overall times for the different conditions. Unfortunately this was not considered prior to the study, so the results in Figures 7, 8 and 9 have a bias due to the initial configuration of the rooms.

4.1.3 Analysis of words used

Participants in the audio-only condition required nearly 2.5 times as many words than the mouse or gesture-based conditions (see Figure 11). A typical example of utterances used by HOG table participants in the audio-only condition is:

> "Behind you. Left. Left. Left. Left. Yes!"

It was common for HOG table participants in the audio-only condition to repeat a direction until the door was in front of the immersed participant. For the mouse and gestures condition HOG table participants seemed comfortable providing a single direction and allowing the immersed participant to continuously rotate until they saw the hand in front of a door. The HOG table participant would see the immersed participant heading towards the correct door and then provide audible confirmation, for example:

> "Left. Yep!"

With the visual cue-based approaches there is no need for the constant confirmation as it seems to be understood that as the immersed participant you just look

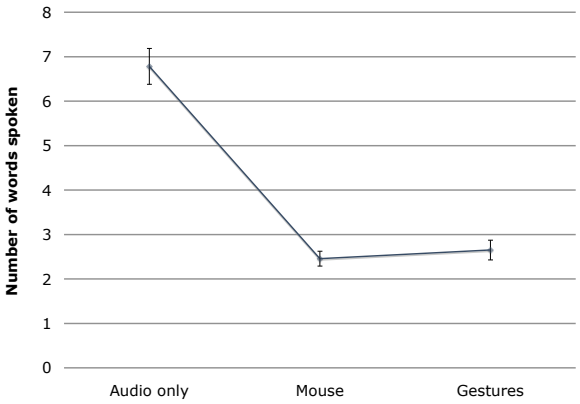

Figure 11: Average number of words per condition.

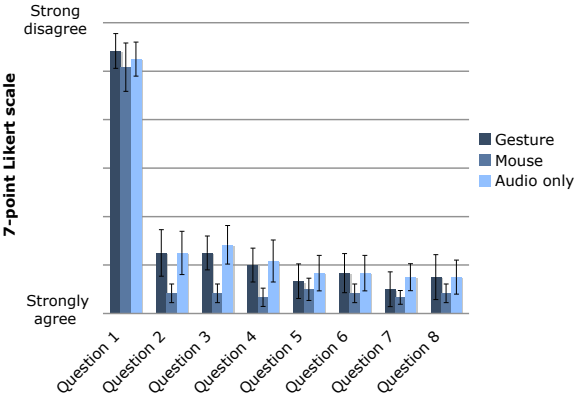

Figure 12: Immersed participants' responses to the post task questionnaire.

around in the direction you are told until you see the hand. Once you see the hand you have found the exit.

In Figure 11 the standard error for the audio-only condition is noticeably larger than the other two conditions. This is attributable to the different approaches the HOG table participants took for the audio-only condition. While some used a single direction technique as described for the gesture conditions above, others seemed to continuously provide directions until the immersed participant had reached the exit, for example:

> "To your right about three doors, keep going, keep going, keep going, keep going, straight ahead, yep keep going, that's it!"

For the other two conditions HOG table participants seemed more likely to restrict their descriptions to a single direction as discussed above.

4.1.4 Questionnaire results

At the completion of the study immersed participants completed a series of questions (see Section 3) comparing the three conditions on a 7-point Likert scale. The mouse conditions consistently ranked lower than the other conditions (see Figure 12). Apart from question 1 (a negatively phrased question) these results suggest that participants thought the mouse made task completion easier because it is best for conveying intention. Results for questions 2, 3 and 4, namely "I always understood my partners directions", "I always knew which door to go through"

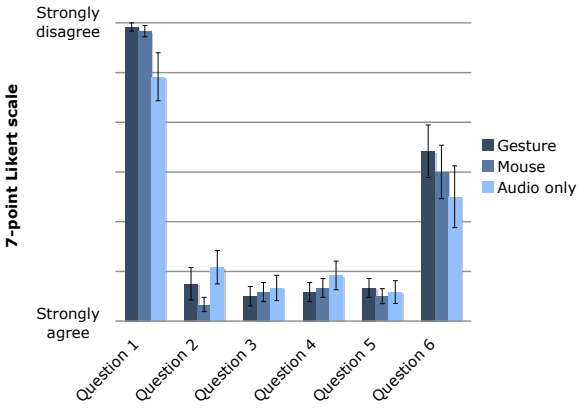

Figure 13: HOG table participants' responses to the post task questionnaire.

and "It was easy for my partner to grab my attention", show a greater difference between the mouse condition and the gesture condition. Responses from the post task interview suggest that the reason was due to the difference in quality of the reconstruction between the mouse and hand gestures:

> "With the mouse the hand reconstruction is stable. So it is already there and you know what it looks like. The orientation is always fixed so that after a while you get used to it which makes it faster."

> "[The camera] was picking up other bits and pieces. Occasionally I would see, not a hand, but something dark and I would go, "oh, what's that over there?" and I would be looking the wrong way."

HOG table participants completed a different set of questions comparing the same three conditions on a 7-point Likert scale. The responses (see Figure 13) are varied compared to the immersed participants responses with no condition clearly outperforming the other. Question 6 ("I had to concentrate very hard to do the task") stands out because the responses were considerably varied and the difference between the gesture (3.42) and mouse (3.0) conditions is the greatest than for any other HOG table participant question. The results suggests that in the gesture condition HOG table participants required less cognitive effort than in the mouse condition, and less cognitive effort again than in the audio condition. This indicates that it is easier to navigate the immersed participant to the exit using the hand gestures than it is to use a mouse or just through speaking. This supports the second hypothesis, however there is not enough data to prove this statistically significant.

While VR participants typically found the mouse interface more effective for the task, it was more common for them to feel a stronger connection with the HOG table participant when using gestures. In the post task interview one participant described the experience as:

> "It looks more real in my opinion. You really think someone is pointing down where as with the mouse it is like on a computer like a cursor or something."

4.1.5 Fatigue

Figure 14 shows the fatigue experienced on various parts of the body for the different conditions. These results were obtained through the questionnaire and HOG table participants were asked to rank the fatigue experienced on the specified parts of the body on a 7-point Likert scale. As the maximum value of the Likert scale is 6 it is clear that there was little discomfort experienced overall for any of the conditions as the highest average discomfort was just 1.92. The audio-only condition produced the least discomfort followed by the mouse condition then the gesture condition. This difference is likely due to the blue perimeter around the table which participants reach over (see Figure 15) evidence of this can be seen in the relatively high discomfort experience in the arm for the gesture condition.

These results are overlaid on an image of a person using the HOG table (see Figure 15). It is clear that there is discomfort associated with the parts of the body involved in reaching over the blue perimeter. Perhaps a reason there is not more discomfort associated with the back and shoulder is that many participants would rest one arm on the top of the blue perimeter and reach down with the other arm.

4.1.6 Monitor use

HOG table participants did not use the monitor of the immersed user view as much as expected. Many participants did not start using it until half way through the task, at which time it was common to realise that what they were doing was not the best approach for the immersed user. In the post task interview one user commented,

> "When I did look at the screen I realised that the finger actually occludes the door and so I thought that it would be better to have the finger above or beside it, so I was trying to position it."

Another said:

> "I hadn't been paying enough attention to the screen so, I think you were doing a good job flying through the hand, because I was putting the hand smack bang on the door and I realise only a fair way into it that it completely occluded the door ... if you are able to represent that some way actually in the tank at the same time it would be much more intuitive because you would not only be able to see what you are positioning but what your partner is seeing at the same time."

5 Follow-up study

A follow-up study was conducted to determine the role of audio between the gesture condition and the mouse condition. The experiment was conducted in the same manner as previously described except participants were not permitted to talk and therefore there was no audio-only condition. Twelve participants from the first study were randomly chosen to participate.

Total time for the mouse condition increased 0.36 seconds over the mouse + audio condition (see Figure 16). This increase predominantly came from the travel time. Total task time for the gesture condition drop an average of 0.88 seconds (see Figure 16). Both the locate time and the travel time were quicker than with audio. This is unexpected as common sense suggests the audio cue such as telling the immersed participant that the door is on the right, starts the immersed participant moving in the necessary direction while the hand can be put into the scene. However it

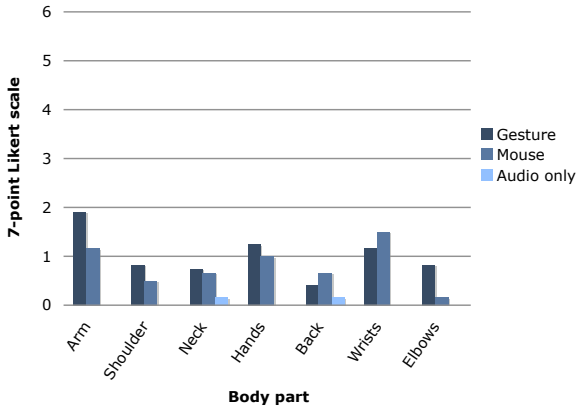

Figure 14: Overview of pain experience in parts of the body while using the HOG table under the various conditions.

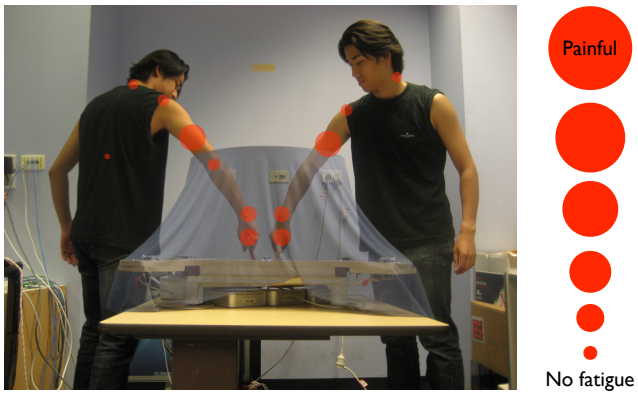

Figure 15: Image that describes the pain on the parts of the body, while a participant is using the HOG table.

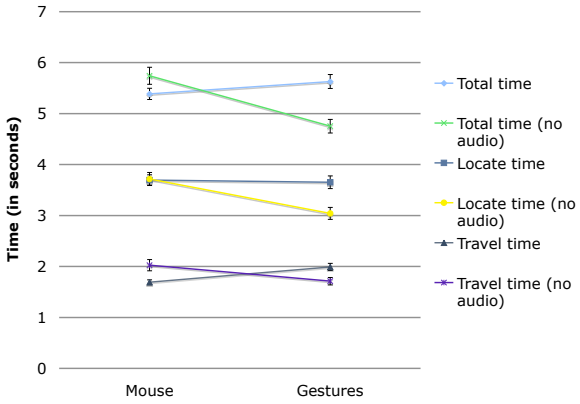

Figure 16: Compares the audio data with the no audio data.

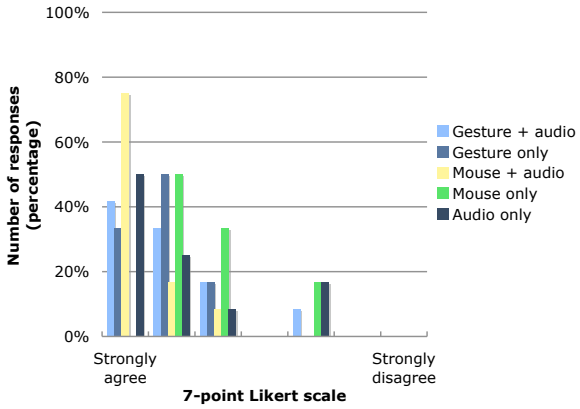

Figure 17: The immersed participants' view on the ease of attention grabbing.

appears that restricting the HOG table participants to stay silent forced them to try alternate ways to gain the immersed participants' attention. The post task questionnaire indicates that HOG table participants were better able to grab the immersed participants attentions. When asked to rate the following statement on 7-point Likert scale "It was easy for my partner to grab my attention" immersed participants responded with an average of 0.83 for the gesture-only condition, this is down from 1.0 when audio could be used (see Figure 17).

The post task interview reveals the HOG table participants' approaches to gaining the immersed participants attention. The first is by pointing to the left or right. In both the previous study and this study HOG table participants were shown that it is possible to point in a direction for the immersed participant, however issuing a direction through voice was the preferred approach. However without the ability to communicate verbally more HOG table participants used this technique. This approach seemed to work well for the immersed participant:

> "For that particular task the hand gestures were actually kind of cool. Once you can kind of see and everything is at the right distance you can understand that that's a finger pointing and it's very obvious, you know what you are meant to be doing. Even though I had no idea that was what you were going to do."

Immersed participants found a way to use the gestures that was not discovered in the previous study using audio communication. Some of the immersed participants discovered that if they looked up they could see the point where the arm first appears in the virtual world. They found they could follow the arm down to the hand to find the door that the HOG table participant was pointing to.

> "If I look up I can see where it goes. So even if it was behind me I could see the rest of my partners arm and so I knew that oh it must be in that direction."

HOG table participants indicated that the task without audio required less cognitive effort. To the statement "I had to concentrate very hard to do the task" HOG table participants responded on average 4.3 on a 7 point Likert scale for the gesture condition where as the average response was 3.4 for the gesture and audio condition. Similarly HOG table participants responded on average 4.3 for the mouse condition and 3.0 for the mouse and audio condition (see Figure 18).

6 Conclusion

The results presented in this paper prove our first hypothesis that visual cue-based approaches are significantly faster for navigation tasks than an audio-only approach for mixed-space collaborative navigation.

While the gesture-based navigation produced similar navigation times to the mouse based-navigation,

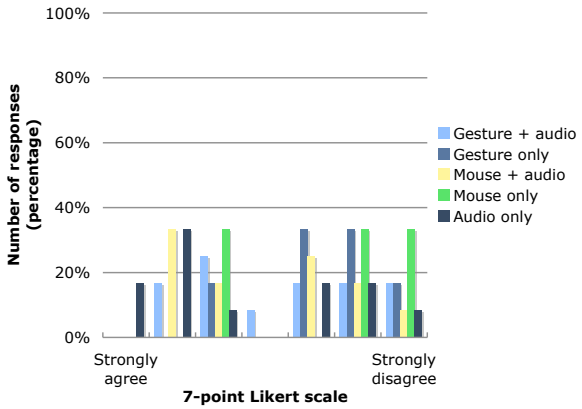

Figure 18: HOG table participants' responses to the question "I had to concentrate very hard to do the task" where "Strongly agree" was given a weighting of 0 and "Strongly disagree" was given a weighting of 6".

there is evidence to suggest that the gesture-based approach requires less cognitive load. HOG table participants commonly believed that the gesture condition was the quickest overall, even though the task completion times are no faster than for the mouse-based condition. This supports the second hypothesis, however there is not enough data to prove this conclusively.

During the gesture conditions a number of different techniques emerged. Some participants pointed directly to the exit and then used verbal commands, others chose to make the immersed participant follow the hand from the starting orientation to the exit, some participants pointed to the left and to the right as a way to describe which way to start heading. To indicate the exit most participants would point directly in front of the door, however, some found that they could point to either side of it so as to not occlude it. The various techniques that emerged from the study highlight the expressive nature of the gesture interface that leverages peoples natural ability to express themselves through gesturing.

References

Billinghurst, M., Kato, H. & Poupyrev, I. (2001), The magicbook: Transitioning between reality and virtuality, *in* 'CHI '01: extended abstracts on Human factors in computing systems', Seattle, WA, pp. 25–26.

Bowman, D. A., Kruijff, E., La Viola, J. J. & Poupyrev, I. (2004), *3D User Interfaces: Theory and Practice*, Redwood City, CA.

Brown, B., MacColl, I., Chalmers, M., Galani, A., Randell, C. & Steed, A. (2003), Lessons from the lighthouse: Collaboration in a shared mixed reality system, *in* 'CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems', Ft. Lauderdale, FL, pp. 577–584.

Darken, R. P. & Sibert, J. L. (1996), Wayfinding strategies and behaviors in large virtual worlds, *in* 'CHI '96: Proceedings of the SIGCHI conference on Human factors in computing systems', Vancouver, British Columbia, Canada, pp. 142–149.

Evans, G. W., Skorpanich, M. A., Gärling, T., Bryant, K. J. & Bresolin, B. (1984), 'The effects of pathway configuration, landmarks and stress on environmental cognition.', *Environmental Psychology.* **3**, 323–335.

Grasset, R., Lamb, P. & Billinghurst, M. (2005), Evaluation of mixed-space collaboration, *in* 'ISMAR '05: 4th IEEE/ACM International Symposium on Mixed and Augmented Reality', Vienna, Austria, pp. 90–99.

Holm, R., Stauder, E., Wagner, R., Priglinger, M. & Volkert, J. (2002), A combined immersive and desktop authoring tool for virtual environments, *in* 'VR '02: IEEE Virtual Reality Conference', Orlando, FL, pp. 93–100.

Kopper, R., Ni, T., Bowman, D. A. & Pinho, M. (2006), Design and evaluation of navigation techniques for multiscale virtual environments, *in* 'VR '06: IEEE Virtual Reality Conference', Alexandria, VA, pp. 175–182.

Leigh, J. & Johnson, A. E. (1996), 'Supporting transcontinental collaborative work in persistent virtual environments', *IEEE Computer Graphics and Applications* **16**(4), 47–51.

Nakanishi, H., Koizumi, S., Ishida, T. & Ito, H. (2004), Transcedent communication: Location-based guidance for largescale public spaces, *in* 'CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems', Vienna, Austria, pp. 655–662.

Poupyrev, I., Weghorst, S., Billinghurst, M. & Ichikawa, T. (1998), 'Egocentric object manipulation in virtual environments: Empirical evaluation of interaction techniques', *Computer Graphics Forum* **17**(3), 41–52.

Schafer, W. A. & Bowman, D. A. (2004), 'Evaluating the effects of frame of reference on spatial collaboration using desktop collaborative virtual environments', *Virtual Reality* **7**(3-4), 164–174.

Schmalstieg, D. & Hesina, G. (2002), Distributed applications for collaborative augmented reality, *in* 'VR '02: IEEE Virtual Reality Conference', Orlando, FL, pp. 59–66.

Stafford, A., Piekarski, W. & Thomas, B. H. (2006), Implementation of god-like interaction techniques for supporting collaboration between outdoor ar and indoor tabletop users, *in* 'ISMAR '06: 5th IEEE/ACM International Symposium on Mixed and Augmented Reality', Santa Barbara, CA, pp. 165–172.

Stoakley, R., Conway, M. J. & Pausch, R. (1995), Virtual reality on a WIM: Interactive worlds in miniature, *in* 'CHI '95: Proceedings of the SIGCHI conference on Human factors in computing systems', Denver, CO, pp. 265–272.

Tlauka, M. & Wilson, P. N. (1994), 'The effects of landmarks on route-learning in a computer-simulated environment', *Environmental Psychology* **14**, 305–313.

Vinson, N. G. (1999), Design guidelines for landmarks to support navigation in virtual environments, *in* 'CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems', Pittsburgh, PA, pp. 278–285.

Yang, H. & Olson, G. M. (2002), Exploring collaborative navigation: the effect of perspectives on group performance, *in* 'CVE '02: Proceedings of the 4th international conference on Collaborative virtual environments', Bonn, Germany, pp. 135–142.