Efficiency of techniques for mixed-space collaborative navigation

Aaron Stafford* Bruce H. Thomas† Wayne Piekarski‡

Wearable Computer Lab
University of South Australia

ABSTRACT

This paper describes the results of a study conducted to determine the efficiency of visual cues for a collaborative navigation task in a mixed-space environment. The task 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 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 results show that visual cue-based collaborative navigation techniques are significantly more efficient than an audio-only technique.

Index Terms: H.5.2 [Information Interfaces and Presentation]: User Interfaces—Evaluation/methodology; I.3.6 [Computer Graphics]: Methodology and Techniques—Interaction techniques; I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual reality

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 [1]. Collaborative navigation is required in situations where navigational targets or conditions of the surrounding environment are constantly changing and therefore difficult to predicted or represent by a computer system.

Frame of reference affects collaboration in a variety of ways [3]. In order to help to better understand interactions in different types of environments, Poupyrev et al. [2] 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.

Previous research has demonstrated the effectiveness of collaborative mixed-space navigation to improve navigation times [1]. 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.

*e-mail: aaron.stafford@unisa.edu.au
†e-mail: bruce.thomas@unisa.edu.au
‡e-mail: wayne.piekarski@unisa.edu.au

In 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 [4]. The 3D reconstruction tabletop (HOG table) 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 a study conducted to determine the efficiency of visual cues for mixed-space collaborative navigation.

2 EXPERIMENT DESIGN

The HOG table communicates over 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 I-glasses HMD that has a horizontal FOV of approximately 26 degrees. A Polhemus 3Space Fastrak magnetic tracker tracks the position and orientation of a HMD and a hand controller with 6 degrees of freedom.

The task was conducted in pairs. One participant was fully immersed in a virtual room. The virtual room is 20m in diameter. 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 is achieved via a button on the hand controller flying the user forward along their current viewing vector.

The other participant (HOG table participant) saw a top-down view of the virtual room the immersed participant was in. 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.

A within subjects study was conducted with a navigation task 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 . 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. In the gesture-based case the reconstruction was updated 7 times a second.

HOG table and immersed participants were located in the same room and therefore able to communicate audibly for all of the conditions. A monitor was positioned close to the HOG table so HOG table participants could see the immersed participants' view. The time between the start of a task and when the immersed user started to move was recorded as the "locate time" and then the time from the locate time to when the immersed user found the exit was recorded as the "travel time" were recorded. Groups experienced the same three conditions in a different order to compensate for learning effects.

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 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.

2.1 Hypothesis

The following hypothesis was formed before conducting the study:

Visual cue-based navigation is more efficient than audio only navigation.

2.2 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.

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). 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 ($p > 0.01$), 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.

The standard error of the total time of the audio only, mouse and gesture conditions were: 0.19s, 0.11s and 0.14s respectively. Clearly there was relatively more variation in the audio only condition than between other conditions. The lack of a visual cue in this condition made it difficult for participants to quickly identify the exit particularly as the number of possible exits increased. There is also more variance in the gestures condition compared to the mouse condition. The larger variance is likely attributable to the variety of approaches that HOG table participants used to direct the immersed participants to the exit. Whereas the HOG table participants in the mouse condition took a more consistent approach.

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 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.

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 a number of 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. Another 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. 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 the exit it so as to not occlude it.

3 Discussion and Conclusion

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

Although there were no significant differences between the efficiency of the gesture-based and mouse-based techniques, gesture-based condition participants were observed using a wider range of techniques than the mouse-based participants for navigating the immersed user. This suggests that the gesture-based approach is a more expressive technique that is better able to facilitate communication of intention.

Further analysis of the results will attempt to extract more information about the expressive nature of the interface. Further analysis of the use of words between the various conditions, analysis of number of possible exits and comfort levels between various interfaces will also be examined.

References

[1] R. Grasset, P. Lamb, and M. Billinghurst. Evaluation of mixed-space collaboration. In *ISMAR '05: 4th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 90–99, Vienna, Austria, Oct 2005.

[2] I. Poupyrev, S. Weghorst, M. Billinghurst, and T. Ichikawa. Egocentric object manipulation in virtual environments: Empirical evaluation of interaction techniques. *Computer Graphics Forum*, 17(3):41–52, Sep 1998.

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

[4] A. Stafford, W. Piekarski, and B. H. Thomas. 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*, pages 165–172, Santa Barbara, CA, Oct 2006.