Exploring Collaborative Navigation: the Effect of Perspectives on Group Performance

Huahai Yang
School of Information
University of Michigan
Ann Arbor, MI 48109, USA
+1 734 913-8097
huahaiy@umich.edu

Gary M. Olson
School of Information
University of Michigan
Ann Arbor, MI 48109, USA
+1 734 647-3576
gmo@umich.edu

ABSTRACT

In this paper, we describe a collaborative navigation task in CVE. As a work in process, we present a process model of the task and design an experiment to test hypotheses generated by this process model. Using this experimental approach, we investigated the effect of the dimension of egocentric-exocentric perspectives on collaborative navigation performance. Results favor an egocentric perspective display. We also discuss the implications of this work for the design of interaction techniques to support collaborative navigation and awareness in CVE.

Categories and Subject Descriptors

H.1.2 [**Models and Principles**]: User/Machine Systems---*Human information processing*; H.5.3 [**Information Interfaces and Presentation**]: Group and Organization Interfaces---*Computer-supported cooperative work, Evaluation/methodology, Synchronous interaction, Theory and models.*

General Terms

Human Factors, Experimentation, Design.

Keywords

Navigation, egocentric-exocentric perspectives, awareness, mental model, user studies, CVE.

1. INTRODUCTION

Many Collaborative Virtual Environments represent such a large space relative to the avatar size that substantial navigation is needed to use the system effectively. As multiple users populate the large space, it is not a trivial task for them to coordinate among themselves and achieve their navigational goal in such an environment [12, 19]. Depending on the work desired and the virtual environment the participants are working in, different types of navigational tasks are required, for which several taxonomies and/or frameworks have previously been proposed [13, 27, 30]. In this work, we focus on the collaborative aspects of navigation in CVE, and describe one type of task called collaborative navigation that many users of CVE find themselves trying to do but failing to do effectively. Built on previous research, we constructed a process model for this type of task, and developed an experimental paradigm to test hypotheses generated by this model. Finally, we present preliminary yet interesting results from our experiment investigating the effect of different perspective displays on group navigation performance. We argue that these findings warrant further exploration in both methodology and design space of supporting collaborative navigation.

2. RELATED WORK

2.1 Collaboration in CVE

In order to support collaborative work in CVE, some issues that characterize such work must be addressed in the design of CVE [7]. There is a basic need to support individualized views, such as those independently controlled viewpoints through which different users can inspect the virtual world from different angles and positions. In Subjective VR-VIBE [26], different users can see viewer-dependent visual representations of the virtual environment. However, taking subjectivity to the extreme would hinder collaboration since a shared context might be destroyed. It is therefore important to support the transition between individual work and group activities [10]. A prerequisite of moving from individual work to group activities is an awareness of other people's activities in the environment.

Research in CSCW has developed an array of techniques to support awareness in 2D workspace and to resolve the often-competing requirements between individual and group work [10]. In 3D virtual environments, techniques using similar approaches have attempted to address the user awareness problem [9]. Many of the techniques center on the concept of an avatar, the embodiment of a user in the virtual environment [2], and provide various enhancements to the avatar such as a view cone, a nose ray or a headlight. Another type of enhancement is the participant list, where another participant's avatar or WYSIWIS (What You See Is What I See) view is shown graphically, and the orientation of the avatar or the view is updated as the participant navigates in the virtual environment. Users can click on the list to switch between the avatar and the view. Essentially, this provides the option of switching between egocentric (first person) perspective and exocentric (third person) perspective views of other participants' viewpoints. However, empirical investigation of the effectiveness of these techniques is still lacking. As indicated in the following section, the egocentric-exocentric dimension of perspective displays is a complex variable that has profound impact on the navigation performance.

2.2 Perspectives

In the field of aviation research, where most available empirical data concerning the effect of perspective displays on navigation come from, the issue of perspective displays is often referred as the issue of Frame of Reference (FOR) [29, 30]. The degree of egocentrism in all FORs is an almost continuous variable that goes from the minimum as in a north-up map to the maximum as in an immersed first person perspective. There are many variations in-between that vary both on lateral referencing, where the view can be immersed, tethered or fixed, and on vertical referencing, where the view can be slaved or fixed on pitch and/or altitude. To tackle the complexity of these different perspectives, we analyze the underlying mental, motor and communication processes needed to accomplish certain navigational tasks given a specific perspective display, and then we pick those representative perspective displays (Figure 2, 3, 4) to verify our understanding of these processes.

3. COLLABORATIVE NAVIGATION

We are interested in supporting tightly coupled synchronized collaboration [17] in CVE. The type of task we are interested in is often required in the so-called Populated Information Terrains (PITs) [3], where user embodiments populate a 3D space that is mapped from some data set. Usually, these data have no intrinsic geometric shape, thus have no natural agreed-upon frame of references such as gravity or latitude-longitude. In the context of a scientific collaboratory [18], we envision such a collaborative scientific visualization application, in which geographically distributed scientists explore a shared 3D space mapped from a large data set. Each of the scientists moves about the space and thus controls her viewpoint independently, and tries to find scientifically interesting trend or points in the data space. They may take a divide-and-conquer approach, partition the space and assign individual to each subspace; or they may each take a different role, and be responsible for finding a specific interesting pattern. No matter what they do, they need to coordinate themselves to accomplish their information-seeking goal, which is usually achieved through navigation in the space. We name this type of collaborative spatial information-seeking activity *Collaborative Navigation*. It usually has the following characteristics or task requirements.

a. Each participant has independently controlled view of the environment. Individual views can differ in term of perspectives, Level Of Details (LOD), or viewer-dependent features. Only with independent views, we can best exploit the potential benefit of collaborative work. [26]

b. There is a need for participants to converge to a common location. This usually happens when one of the participants finds something interesting and wants others to see it. It is often accompanied by words like "Come here! Look at this." And this sentence could imply two different requests: a) looking at a point of interest, where the listener needs to face the direction where the speaker is looking at to complete this request; and b) looking from a particular orientation and position, where the listener needs to navigate to the same location where the speaker is to complete the request. These two situations entail different design strategies to support them.

c. It is beneficial to understand others' perspectives. As our experiment will show, there is a large communication overhead between participants if they do not know what their partner is looking at, and do not understand how the other's viewpoint is related to their own. Thus it is essential for CVE designers to support awareness of the others' perspectives. Although from the bulk of psychological literature on spatial perspective taking [8, 21, 22], we know it is rather difficult for people to image perspectives other than their own, it is still possible to facilitate perspective taking through proper exposure to multiple perspectives [23, 28].

d. Participants have to know the environment to some extent. Here, "know" implies the ability to remember or recognize certain spatial features of the environment, such as overall shape of the object distribution. In fact, some virtual environment systems have environment learning as one of their goals [4]. For a scientific visualization application of CVE, we believe that environment learning could prove to be a very important factor for successful adoption of the system.

Given these task requirements, we narrow down the scope of mental and communicational processes involved in doing this type of task. Thus it is possible to sketch out a framework for a process model of collaborative navigation in a large-scale 3D space. This allows us to develop an experimental paradigm to test and modify the model, and it also fuels our further exploration of the design space for supporting collaborative navigation.

3.1 Process model

Spence proposed a general framework for navigation process [27]. His model spans a whole range of navigation tasks including those in the physical, virtual and information space. In fact, his definition of navigation is "the creation and interpretation of an internal mental model". In contrast, Wickens' model of navigation in 3D space [30] is more detailed, more specific for spatial navigation, and involved several closed-loop processes. However, his model leaves no place for an internalized model of the environment such as a cognitive map, except for an estimate of the current location vector. Situation awareness, his term for an awareness of the geographical surroundings, is only a by-product of the navigation process, but not an integral part of the process.

However different, the above frameworks for navigation bear significant similarity. They all have the common components of goal, strategy, search (scan, browse) and act (control), and consist of recursive loops. In developing our process model of the collaborative navigation process, we take these components into account. Nevertheless, one of the foci of our effort is to understand the collaborative aspects of the task. To do so, we turn to research on spatial language communication for inspiration. Specifically, we are interested in the behavior of referring in conversation, since many of the difficulties in collaborative navigation are related to the problem of understanding what others are referring to [12].

In this area, Clark's model of referring communication [6] is representative. In his model, referring is a recursive process of seeking mutual acceptance between conversation partners. A typical acceptance process involves the following: one person presents a noun phrase to initiate a reference, her partner then evaluates it, depending on the outcome, the noun phrase is either accepted or refashioned (repaired, expanded, or replaced) until a mutual acceptance is achieved. This model can make some predications about some communication phenomena. In CVE, however, a pure verbal conversational model is not sufficient; since participants could use graphically represented action instead of utterances to communicate. These actions often serve as an indication or verification of the referent. For example, a user may manipulate her avatar to face towards the referred object, or highlight the object using a light ray while saying “This one?”.

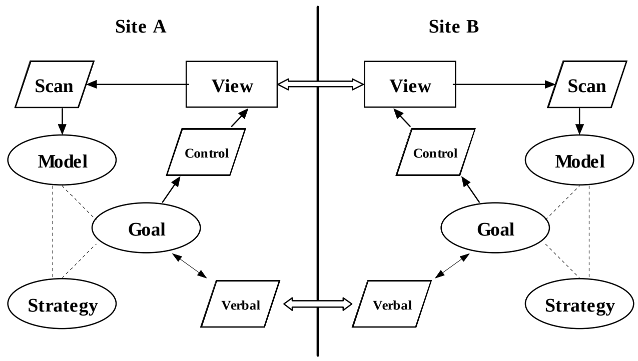

Figure 1 Process model of collaborative navigation

Based on previous research, our current model, shown in Figure 1, is a work in progress. The centerpiece of our model is a Goal stack, which holds the user’s task goal as well as the sub-goals generated along the way of trying to achieve the task goal. Since navigation, as a type of information seeking behavior, constantly generates and solves smaller goals in order to reach a given goal [1], a LIFO stack is needed to keep track of these goals. The dashed-line loop is an Evaluation process for handling incoming inputs from the visual or verbal channel, and for deciding which operation will be taken on the goal stack. Aside from self-initiated goals, verbal input and visual scan input could trigger the evaluation process, and consequently some new goals may be pushed in, or some goals be popped out. A goal can be popped out of the goal stack because of its evaluation being complete, or too difficult.

The solid arrow line from Goal to Model constitute the visual searching and tracking loop [30]. For example, consider a task of looking for an electric outlet to plug in my laptop at a library. First, a Goal of “finding an outlet” is pushed into the goal stack, and the evaluation process starts. From my experience in the library, I realize that the outlet is usually near the floor on the wall, so the Strategy component pushes a Goal of “looking for a wall” into the stack to be evaluated. Since this is a goal that is immediately doable, the Control output component is activated and I start to move my head around. Here, the View I see is physical, but the View in CVE is either synthesized (as in pure VR) or mixed (as in Augmented or Mixed Reality), and determined by the system design. I then visually Scan the environment, and the visual information is integrated into my mental Model of the environment, which is what usually called a cognitive map [15]. The Model is evaluated and it is decided that I am facing a wall, thus the Goal on the top of stack is considered achieved and popped out of the stack. A new Goal of “looking along the downside of the wall” is pushed in, or it could have already been there from the last loop, depending on how the Strategy component works. Now, a new round of control-scan-evaluation loop starts. The process goes on until I achieve the original goal of “finding the outlet”.

We have discussed the individual navigation process in our model; the collaborative part comes in when two individual navigation processes interact. The current state of technology supports two types of interaction between geographically distributed individuals: verbal communication and correlated views. Verbal communication is easy to set up and is often preferred. However, it is a serial process and time-consuming. Naturally researchers turn to visual channel, and try to make views across different sites correlated so that information can be transferred. The simplest form of correlated view is WYSIWIS such as those found in SharedX. However, as mentioned before, an individualized view is also important for collaborative work, and the struggle lies in how to balance an individualized view and collaborative work. We believe what is needed for a successful collaboration is not the shared view, but the shared mental model of what being worked on. Because of the currently limited channels (visual, verbal) we have available to convey information across distance, we have to better design the information presentation in these channels, so that the collaborative partners can correctly infer a shared model of the work environment. To do so, we need to examine the bigger loop across the site boundary: evaluation-verbal-evaluation-control-view-scan. And the goal is to make the individual models gained during these big loops compatible with each other.

The verbal and visual channels of communication need to supply information that is lacking in the other channel but is important for inference of a shared mental model. For example, the talking-head type of videoconference is found to provide no extra value compared to audio only conference [5]. The reason, we would argue, is that it supplies no extra information to help infer a shared mental model of what being worked on. Hence it is not surprising that some video systems that focused on the working objects proved to be much more successful [16]. Here, we are interested in finding the proper correlated views that ease the inference of a shared mental model. Ideally, less verbal communication would be needed before a user can know (meaning that information has been integrated into part of her mental model) what her partner is looking at. Views in different sites can be correlated in many ways, such as perspectives, LODs, scales, appearance, and so on. Given a task, we want to study what are the good ways to correlate views in different sites; such that more useful information for constructing shared mental models is transferred. And the amount of verbal communication is needed to compensate for the lack of such information would be a measure of how good these correlated views are. We now introduce an experimental paradigm to do this kind of investigation.

3.2 Experimental paradigm

Similar to many researches on spatial language communication, especially those studying referring communication [6, 14, 24], we assign roles to experiment participants so that the process we intend to study can be observed in a formal fashion. The task is set up in such a way that one participant possesses certain information that must be conveyed to another participant for the task to be completed. In our experiment, the task was a collaborative exploration task in a virtual water tank, and participants drove a virtual submarine to find some targets in the tank. Only one of the targets was flashing, and the submarine needed to hit the flashing one. After the flashing target was hit, it stopped flashing, and another one started flashing. One participant took the role of a Driver, who has control of the submarine, but in her view the target was not flashing. Another participant was a Guider, who could see the target flashing if the target was in her view, but did not have control of the submarine. Thus, the guider had somehow to convey to the driver the information about which target is flashing. This could be done either through verbal communication or correlated views. What we were interested in is how different techniques of correlating views affect the overall collaboration performance, in term of quality of the collaboration and the time spent doing it. As we have already mentioned, another useful indication of the quality of correlated views is the amount of verbal communication needed to repair and compensate for the lack of information.

Using this experiment paradigm, we examined one way of correlating different users' views - showing other person's view from different perspectives. As we mentioned before, the egocentric-exocentric dimension is the main factor governing the variations on perspectives. Here, we wanted to know at which point along this dimension, the transfer of missing information from the Guider to the Driver is best supported under such a perspective display of the Driver's view. That is to say, we varied the Guider's views as our independent variable, and under different conditions, the Guider looked at the Driver's view from different perspective. For example, one of the most exocentric perspectives was to look at the submarine that the Driver was driving from a fixed position in the tank (Figure 4). The Driver always used a normal driving perspective, i.e. a first-person perspective. The following sections describe the experiment in detail and discuss the results.

4. EXPERIMENT

Based on our process model of collaborative navigation, we developed some hypotheses regarding the effects of varying perspectives on collaborative navigation performance for the above experiment. In this experiment design, the Guider's *control* components is removed, thus navigation *goals* can only be achieved through the cross-sites big loop. The difference in the Guider's perspectives directly impacts the *view*-to-*view* link, which in turn affects the *model*s of both sites.

We considered the overall goal of hitting the flashing target as comprised of two major sub-goals. The first sub-goal is to *search* for the flashing target; the second is to *travel* to the flashing target once it has been found. Of course, each of these sub-goals could be furthered divided into smaller goals such as "move towards the red wall". The time and effort needed for the subjects to achieve these goals were the metrics for examining the effects of different perspectives. Moreover, since we were concerned with how good participants learned about the virtual environment, we tested their *global judgment* about the target distribution and pattern.

Table 1 summarizes our predictions about how different perspectives might influence the achievement of these sub-goals. Search was predicted to be slow for egocentric perspective because egocentric perspective usually bring less of the world into view given the same geometric field of view (GFOV) [29], hence it would take longer for the guider to scan the scene to find out which target was flashing. Travel was thought to be difficult under exocentric condition because it usually has low resolution due to the large viewing distance, thus making it hard to determine the submarine orientation by visual scanning. In addition, under exocentric perspective, the view orientation may be different from that of the Driver's view, thus more effort is needed to construct the model of the submarine position relative to the flashing target through mental rotation [25]. Under the egocentric condition, the global judgment of target pattern would be bad because the subject had to put together piecemeal visual scenes in order to get the big picture.

Table 1 Hypotheses: the effect of perspectives on navigation

	Search	**Travel**	**Global judgment**
Egocentric perspective	Slow	Easy	Bad
Exocentric perspective	Fast	Difficult	Good

4.1 Method

4.1.1 Subjects

Twenty-four pairs of experienced computer users (46 of them use computer everyday, the rest several times a week) participated in the study. Fourteen were females. All but eight had some 3D application experience through either games or visualization tools. The average age of the participants was 21.7, ranging from 19 to 28 years of age. All participants had normal or corrected-to-normal vision and were paid $20 each for their participation.

4.1.2 Apparatus

The experiment paradigm described in section 3.2 was used for the task. The task was developed in M-SCOPE [31], a synchronized collaborative virtual environment system based on Java3D and JSDT, and run on two 1.4 GHz Pentium4 Dell Precision 340 workstations with 17-inch monitors. GeForce2 graphic cards were used for rendering the scenes under 32bits true color 1024x768 resolution. These two computers were located in two rooms and connected through an Ethernet connection. There was no noticeable frame rate or network lag during the experiment. These two rooms were equipped with video cameras, microphones and speakers, and were connected through audio links. Subjects could talk to each other during collaborative experimental blocks, but they could not see each other.

The virtual water tank was measured 400 meters on each side. Each wall had different colors, with color-matched grids providing position and orientation cues. The submarine was 6.4 meters long, 1 meter in body radius and 2.4 meter in wingspan. Targets were all red balls with a radius of 5 meters, and their positions formed some geometric patterns. Each target ball had a string connecting it to the floor, providing position references. 100 transparent objects were scattered around the tank, and maintained fixed positions throughout the experiment. Each of these objects had a stem pointing towards the floor, providing reference to "up-down" direction. All of the perspective displays in this experiment had a GFOV equals to 90°.

A 3-button Microsoft IntelliMouse was used as input device and achieved 6 degrees of freedom movement in 3D space, as shown in Table 2. According to post-experiment questionnaire, subjects rated these mouse controls relatively “intuitive and easy to learn” (mean rating = 3.61, out of 5).

Table 2 Mouse controls for navigation

Cursor	*Left button*	*Middle button*	*Right button*
Above aimer	Move forward	Pitch up	Ascend
Blow aimer	Move backward	Pitch down	Descend
Left of aimer	Turn left	Clockwise roll	Shift to left
Right of aimer	Turn right	Anti-clockwise roll	Shift to right

4.1.3 Design and Conditions

A repeated within-subject design was used to compare navigation performance between the following four different conditions. All these conditions were applied to the Guider. Each participant of a pair took turns to play roles of Guider or Driver. Half of the subjects took Guider role first, another half Driver first. Two 3x3 Latin Square schemas were used to counterbalance the three collaborative conditions and two roles. However, due to a procedure difficulty, the Single condition was not counterbalanced.

Single condition: This is a condition when subjects drive the submarine to search and hit the flashing target alone**.**

First Person condition**:** This is the collaborative condition when the Guider looked at exactly what the Driver looked at (WYSIWIS), except that the Guider could see target flashing while the Driver could not. It is the most egocentric perspective condition.

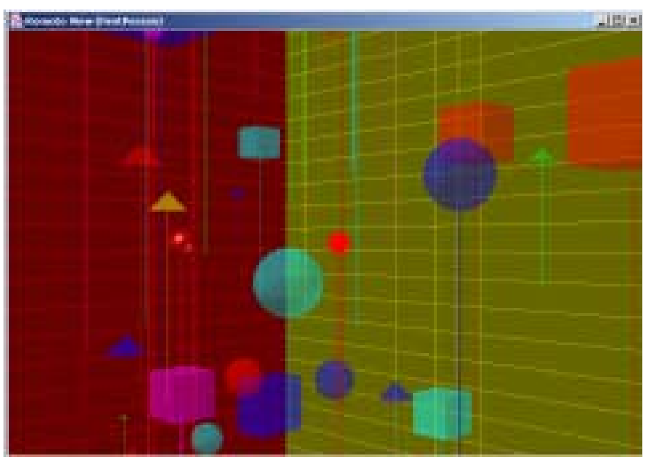

Figure 2 First Person perspective

Tethered: This perspective is in the middle of the ego-exocentric dimension. The view looks at the submarine from right, above, and behind; each has an offset of 150 meters, thus both the azimuth and elevation angle are 45°. As the submarine moves, the view changes both in position and direction to maintain its relative position and orientation to the submarine.

Third Person: This is the most exocentric perspective. The view is stationary, looking at the center of the tank from right, above, and behind; and each has an offset of 480 meters, thus both the azimuth and elevation angle are 45°. All of the targets are in the view all the time. When the Driver drives the submarine around, the Guider can see the submarine move around in the tank. Essentially, it is a 3D perspective map plus a dynamically updated submarine position indicator.

Both Tethered and Third Person condition have the same enhancement done to the submarine, in that 3 white lines are connecting the submarine to the 3 walls on the canonical positive directions of x, y, and z axis. These axis lines provide information about position of the submarine.

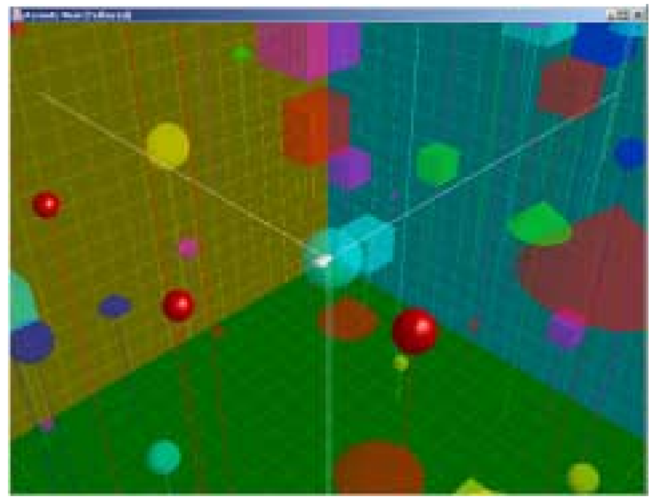

Figure 3 Tethered perspective

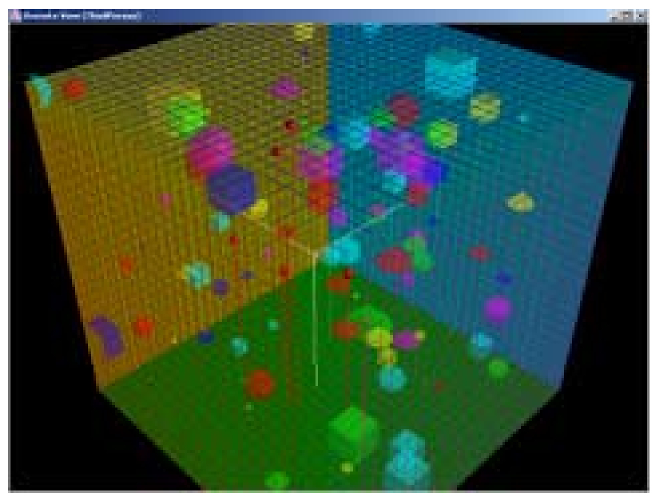

Figure 4 Third Person perspective

4.1.4 Measurements

a. Search Time: A timer is started when a new target starts flashing. When the flashing target’s center is within the Guider’s view frustum, the timer is stopped, and the elapsed time in-between is recorded as search time, i.e.

Search Time = $T_{\text{target appears on guider's screen}} - T_{\text{target starts flashing}}$

This is the time taken for the subjects to find the target.

b. Travel Time: is defined as:

Travel Time = $T_{\text{target collides with submarine}} - T_{\text{target appears on guider's screen}}$

This is the time taken for the target to be hit by the submarine after it has been found is recorded as Travel Time.

c. Global Judgment: After each block of experiment, the 3D scene disappears, and the computers present a multiple-choice question to both the Driver and the Guider, asking about the overall pattern targets formed in that block. They were not allowed to discuss the question. Answers to these questions are recorded automatically.

4.1.5 *Procedure*

The duration of the experiment ranged from 40 min. to one and half hours. It consisted of a series of experiment blocks. First, subjects did 1 tutorial block, in which individual navigation skills were learnt and practiced. Then, they went through a series of testing blocks. Both subjects must get fast enough in order to proceed to collaborative blocks, or another testing block would begin. Based on a pilot test, a passing criteria of *average time of hitting a* target<40 seconds and *maximum time<80* seconds was established. This ensured that each subject reached certain level of competence of navigation in the environment. The last testing block was counted as Single condition. After passing the test, subjects worked on 6 collaborative conditions as described in section 4.1.3.

4.2 Results and Discussion

4.2.1 *Search Time*

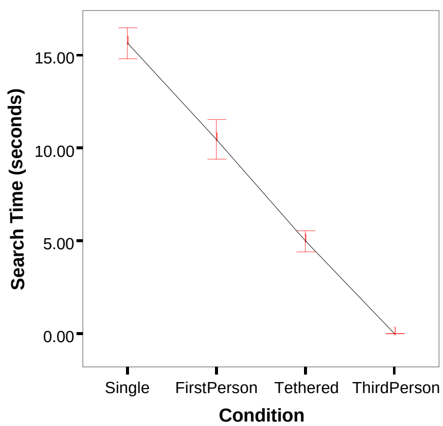

Figure 5 Search Time (error bar show mean +/- 1.0 SE)

As Figure 5 shows, less egocentric perspectives incur faster search. Using Tamhane's T2 multiple comparison (because of unequal variance), all of the pair wise differences are significant (all p=.00). This is predicted by our hypotheses. One interesting thing here is that in the collaborative First Person condition, the Guider used exactly the same perspective as she did in the Single condition, but she managed to get the flashing target on her screen faster than in the Single condition. Remember that she did not have direct control of the view in collaborative conditions. Despite the time delay in talking to her partner to get her view changed, she, along with her partner, still found the target quicker than search alone. Before we jump to any conclusion or explanation, however, we should note that there was a learning confound because collaborative conditions were always done after the Single condition was done. Nevertheless, we believe the observed effect may do exist because the communication time delay was longer than the possible time saving through learning, as subjects had already well learnt through testing blocks.

4.2.2 *Travel Time*

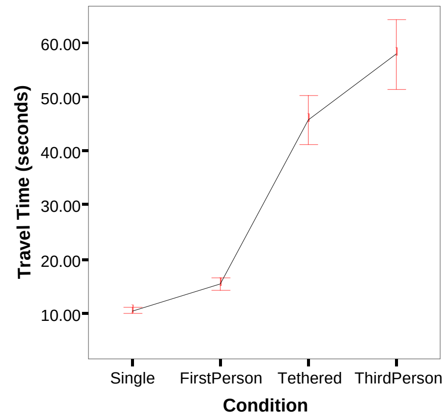

Figure 6 Travel Time (error bar show mean +/- 1.0 SE)

As Figure 6 shows, less egocentric perspectives incur longer travel time. Except for the differences between Tethered and Third Person conditions (p=.56, n.s.), all other pair-wise differences were significant using Tamhane's T2 multiple comparison (all p=.00). Despite the learning confound we just mentioned, in First Person condition, more time is needed than that in Single condition to reach a target after it has been found because of the communication overhead. The huge jump from First Person (mean=15.4sec) to Tethered perspective (mean=45.7sec) is striking. As we observed, for some subjects in the Tethered condition, enormous time was spent on verbal communication, and the Guider seemed to find it very difficult to convey the information as to which target was flashing.

What is the difference between First Person condition and the Tethered condition? One major difference is that there is an offset between the Driver's view orientation and the Guider's in the Tethered condition. Because of this orientation discrepancy, in the bigger loop across site boundary in our process model, the Guider often can not get her expected view change, thus is unable to achieve her sub-goal. She then has to discard her original sub-goal and come up with another goal, i.e. another way to give instruction. For example, the Guider gives a verbal instruction such as "move forward", the Driver then zooms in the view to move the submarine forward, and because of the way Tethered view is constructed, in the Guider's view, the submarine will appear to be moving diagonally up and right. At this point, as we observed, the Guider was often confused and the typical reaction was "no, no, go back to where you were". The Guider then tried to figure out another way to give the instruction. In order to achieve her sub-goal, the Guider must constantly do a mental rotation in constructing her model, i.e. transforming the submarine movement from apparently up and right to straight forward. As a mental rotation of a static image is already hard [25], we have reason to believe it is much harder to do in real-time. Moreover, because the Guider has to speak all the time, from the previous evidence on performance hit due to language – spatial task interference [11], we suspect the difficulty of mental rotation is significantly magnified in a collaborative navigation task.

4.2.3 Global Judgment

Contrary to our hypotheses, we did not find any significant difference between egocentric and exocentric perspectives on Global Judgment score. Since the data distribution of Global Judgment score is close to normality (mean=3.56 out of 8, std.=1.81, Shapiro-Wilk test p=.03), we ruled out the possible explanation of having a floor or ceiling effect. We feel that this measurement reflects more of the individual differences in term of the ability to form a cognitive map by observation of the scene.

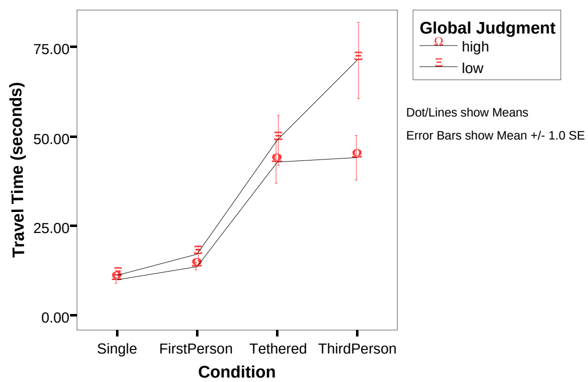

Figure 7 Travel Time by Global Judgment

With this understanding, we divided the subjects (excluding 3 subjects who had incomplete travel data) into two groups according to their Global Judgement scores: group 1 has scores>3, n_1=23, group 2 has scores<=3, n_2=22. The redraw of Travel Time results by this grouping is in Figure 7. For Third Person condition, the half of subjects with high Global Judgment score did significantly better than those with lower Global Judgment score (t_{43}=2.22, p<.05). Overall, this result is consistent with an egocentric/non-egocentric explanation of spatial perspective taking [24]. By this account, what is difficult is producing and comprehending any perspectives other than those of their own, thus speaking non-egocentrically is uniformly hard, regardless of the size of the perspective offset. Our results qualify and extend this reasoning, in that for people with lower spatial ability in forming cognitive map from observation of the scene, the degree of perspective discrepancy does make a difference. The more exocentric the perspective is, the harder it is for people to do mental rotation in order to give correct instructions.

4.2.4 Total Time

Adding the Search Time and Travel Time together, as shown in Figure 8, the Total Time strongly favors an egocentric perspective. This finding was not expected since from the hypotheses, it seemed that exocentric perspective would be better since it excelled in two aspects while only being bad at one aspect. But the travel aspect that exocentric is bad at is a dominating factor of time under collaborative conditions, so that the overall performance favors an egocentric perspective. Under a more feature-rich virtual environment, this strong advantage of egocentric perspective may be reduced, since the predication is that it is easier to give travel instruction if there are more distinct landmarks in the environment to be potentially used for reference, which tends to favor exocentric view. Although we suspect the same trend of results will emerge, we are still considering revising our experiment paradigm to work in a more feature-rich virtual environment.

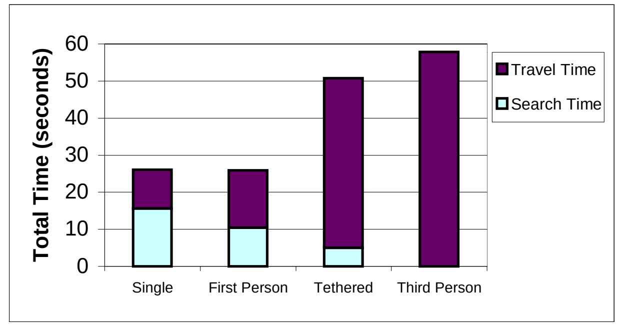

Figure 8 Total Time composition

5. IMPLICATIONS

5.1 Collaborative navigation process

Although it is a work in progress, we have shown that our process model could make some predictions about the collaborative navigation performance under different display conditions. Our experimental paradigm is working, and we believe it can be used to examine other correlated viewing techniques for collaborative navigation. For the research problem of how different perspectives affect collaborative navigation, many interesting issues still remain for further investigation. For example, what will happen if the Driver drives in a Tethered perspective that matches with the Guider's perspective? What will happen if two people each drive their own submarine? Also, we did not control the GFOV so that each perspective brings unequal amount of the world into view [29], it would be interesting to see how that affects the Search Time if we do control GFOV. Another aspect that could be a fertile ground for research is the strategy aspect of collaborative navigation process, and detailed analysis of the process through annotation and coding could prove to be useful here.

5.2 Tool design

We believe the experiment results point us to some directions in exploration of the design space for supporting collaborative navigation. One lesson learnt is that it is harmful to correlate views across sites in a way that requires real-time effortful mental operation such as mental rotation. A WYSIWIS approach seems to be the best way to correlate views. However, as we have seen, that is too restricted and does not fully support the potential of collaborative work. Therefore, we propose an integrated approach to offset the benefit and cost of individual views and shared views. The basic idea is to use animation, by providing a smooth transition between an individual's own view and the others' views, so we can better support transfer of information needed to build a shared mental model. Despite the fact that looking at animations takes time from real work, we believe some overall time saving can be achieved through reduced communication overhead. Another possible advantage of animation is its support of better environment learning. To achieve effective animation under various constraints, some advanced camera control techniques will be needed [20]. Realizing it is not a trivial problem, and we are exploring several techniques through prototyping.

6. ACKNOWLEDGMENTS

We would like to thank Weimin Mou, Yunyao Li, Judith Olson for useful comments; Nathan Bos, Wei Huang, Jing Wang for helping with the experiment setup. The work was supported by NSF Grant IIS-0085951

7. REFERENCES

[1] Bates, M.J. The design of browsing and berry picking techniques for the on-line search interface. Online Review 13(5), 1989, 407-431

[2] Benford, S., Bowers, J., Fahlen, L.E., Greenhalgh, C. and Snowdon, D. User Embodiment in Collaborative Virtual Environments. in Proceedings of ACM CHI'95, 242-249.

[3] Benford, S., Snowdon, D. and Mariani, J. Populated information terrains: first steps. in Earnshaw, R.A., Vince, J.A. and Jones, H. eds. Virtual Reality Applications, Academic Press, New York, 1995, 27-39.

[4] Bliss, J.P., Tidwell, P.D. and Guest, M.A. The effectiveness of virtual reality for administering spatial navigation training to firefighters. Presence - Teleoperators and Virtual Environments, 6(1), 1997, 73-86.

[5] Chapanis, A. Interactive human communication. Scientific American, 232, 1975, 36-42.

[6] Clark, H. and Wilkes-Gibbs, D. Referring as a collaborative process. Cognition, 22, 1986, 1-39.

[7] Churchill, E.F. and Snowdon, D. Collaborative Virtual Environment: An Introductory Review of Issues and Systems. Virtual Reality, 3, 1998, 3-15.

[8] Cox, M.V. Order of the acquisition of perspective-taking skills. Developmental Psychology, 14(4), 1978, 421-422.

[9] Dyck, J. and Gutwin, C. Groupspace: A 3D Workspace Supporting User Awareness. in Proceedings of ACM CHI2002, 502-503.

[10] Gutwin, C. and Greenberg, S. Design for Individuals, Design for Groups: Tradeoffs between Power and Workspace Awareness. in Proceedings of ACM CSCW'98, 207-216.

[11] Hermer-Vazquez, L., Spelke, E.S. and Katsnelson, A.S. Sources of flexibility in human cognition: Dual-task studies of space and language. Cognitive Psychology, 39(1), 1999, 3-36.

[12] Hindmarsh, J., Fraser, M., Heath, C., Benford, S. and Greenhalgh, C. Fragmented Interaction: Establishing Mutual Orientation in Virtual Environments. in Proceedings of ACM CSCW'98, 217-226.

[13] Jul, S. and Furnas, G.W. Navigation in Electronic Worlds: A CHI97 Workshop. ACM SIGCHI Bulletin, 29(4). 44-49.

[14] Keysar, B., Barr, D.J., Balin, J.A. and Brauner, J.S. Taking perspective in conversation: The role of mutual knowledge in comprehension. Psychological Science, 11(1), 2000, 32-38.

[15] Kuipers, B. The Cognitive Map: Could It Have Been Any Other Way? in Pick & Acredolo (Eds.), Spatial Orientation: Theory, Research, and Application, 1983, 345-359.

[16] Nardi, B.A., Schwarz, H., Kuchinsky, A., Leichner, R., Whittaker, S. and Sclabassi, R. Turning Away from Talking Heads: The Use of Video-as-Data in Neurosurgery. in Proceedings of ACM CHI'93, 1993, 327-334.

[17] Olson, J.S. and Teasley, S. Groupware in the Wild: Lessons Learned from a Year of Virtual Collocation. in Proceedings of ACM CSCW'96, 419-427.

[18] Olson, G.M., Atkins, D.E., Clauer, R., Finholt, T.A., Jahanian, F., Killeen, T.L., Prakash, A. and Weymouth, T. The Upper Atmospheric Research Collaboratory (UARC). interactions, 5(3), 1998, 48-55.

[19] Park, K.S., Kapoor, A. and Leigh, J. Lessons Learned from Employing Multiple Perspective in a Collaborative Virtual Environment for Visualizing Scientific Data. in Proceedings of ACM CVE'2000, 73-82.

[20] Philips, C.B., Balder, N.I., Graneri, J. Automatic Viewing Control for 3D Direct Manipulation. Proceedings of 1992 Symposium on Interactive 3D Graphics. 71-74

[21] Piaget, J. and Inhelder, B. The child's conception of space. Routledge & Kegan Paul, London, 1948, 1956.

[22] Rigal, R. Right-left orientation, mental rotation, and perspective-taking: When can children imagine what people see from their own viewpoint? Perceptual & Motor Skills, 83(3, Pt 1), 1996, 831-842.

[23] Schatzow, M.D., Kahane, D.C. and Youniss, J. The effects of movement on perspective taking and the coordination of perspectives. Developmental Psychology, 16, 1980, 582-587.

[24] Schober, M.F. Speakers, addressees, and frames of reference: Whose effort is minimized in conversations about locations? Discourse Processes, 20(2), 1995, 219-247.

[25] Shepard, R. N. & Metzler, J. Mental rotation of three-dimensional objects. Science, 171, 1971, 701-703.

[26] Snowdon, D., Greenhalgh, C. and Benford, S., What You See is Not What I See: subjectivity in virtual environments. in Frameworks for Immersive Virtual Environments (FIVE'95).

[27] Spence, R. A Framework for Navigation. International Journal of Human-Computer Studies, 51(5), 1999, 919-945.

[28] Weatherford, D.L. and Cohen, R. Influence of prior activity on perspective taking. Developmental Psychology, 16(3), 1980, 239-240.

[29] Wickens, C.D. and Prevett, T.T. Exploring the dimensions of egocentricity in aircraft navigation displays. Journal of Experimental Psychology: Applied, 1(2), 1995, 110-135.

[30] Wickens, C.D. Frames of reference for navigation. in Gopher, D. and Koriat, A. eds. Attention and Performance XVII, MIT Press, Cambridge, MA, 1999, 113-144.

[31] Yang, H. Multiple perspectives for collaborative navigation in CVE. in Proceedings of ACM CHI2002 Extended Abstract, 560-561.