IPT & EGVE Workshop (2005)
R. Blach, E. Kjems (Editors)

Reducing Fragmentation in Telecollaboration by Using IPT Interfaces

D. Roberts, M. Al-Liabi, R. Wolff, O. Otto and A. Al-Khalifah[2]

Centre for Virtual Environments, University of Salford, UK
[2] Department of Computer Science, University of Reading, UK

Abstract

Telecommunication systems, such as AccessGrid, allow collaboration across a distributed team. However, these systems typically introduce fragmentation into the view of the shared environment. Many have found that IPT systems offer several important advantages above other display technologies in supporting distance working. This study focuses on fragmentation, which has previously been shown to induce problems in efficient object referencing within a shared virtual environment accessed through desktop displays. We have attempted to repeat the experiment while varying the display type. The results reinforce previous studies by showing a significant improvement in task performance when the entire team uses IPT displays. We further show that the improvement is unlikely to come in this case from more natural interaction or navigation and thus postulate that it arises from more efficient mutual orientation towards objects of interest, arising from a reduction in fragmentation.

Categories and Subject Descriptors (according to ACM CCS): I.3.6 [Computer Graphics]: Methodology and Techniques

1. Introduction

Telecommunication technology allows collaboration across a distributed group of people, offering many advantages in today's globalised socio-economic information culture. Characteristics of display technology can play a major role in the effectiveness of distant teamwork. A key factor is fragmentation of the shared environment.

When collaborating through a traditional telephone, the shared environment can only be perceived through sound and imagination that sound and particularly conversation conjure up. With teleconferencing technologies, such as AccessGrid, people can look into each other's world. They therefore have a shared space, however each has a limited view of the part of that space occupied by the remote collaborator and can not move around it or interact with objects within it. These technologies can also share software applications, although typically only one person can interact with the software at a time. More fundamentally, the team is in no way immersed in the information environment, which limits the naturalness, and we would argue the performance, of communicating attention and interaction. Studies that reinforce this argument from various perspectives include [HFHB, Art96, Kje01, LV02, PWBI98].

Collaborative virtual environments (CVE) allow people to share a space in a fair and spatially unconstrained manner. This helps to reference objects within the space and to indicate focus of attention or activity. A well known study by Hindmarsh et al. [HFH*00], showed that fragmentation was still a problem within CVEs and linked this to the low field of view offered by desktop interfaces. The study demonstrated that limited field of view made it difficult to assess what a remote participant was aware of, and slow and unnatural control of this limited field of view through a desktop interface meant that the environment was observed through "fragments" making it difficult to reference objects. It was suggested that it is seldom possible to observe both the remote collaborator and the object to which he points at the same time.

Following the direction of a reference requires slow control of ones own field of view, and this turns referencing into a task in itself rather than simply a tool. The work concluded that the straightforward translation of human physical embodiments into CVEs, however detailed, are likely to be unsuccessful unless participants can also be provided with the perceptual capabilities of physical human bodies, including a very wide field of view and rapid gaze and body movement.

We believe that IPT presently comes closer to meeting this requirement than any other display type. Several recent studies have used IPTs to interface to CVEs and have found them to be very effective [SSA*01, RWO*04, HSS*05]. We are part of a growing community that believe networked IPTs bring us considerably closer to resembling a face-to-face meeting between a distributed group. A growing wealth of research adds weight to this argument.

When compared to desktop displays, linked IPTs have been shown to improve capabilities [RWO*04,HSS*05], impact on role, increase feelings of contribution and collaboration and increase task performance. We suspect that these improvements come from a set of factors that together allow people to consciously and subconsciously use their body in a natural way to observe and interact with the environment and avatars within it.

In theoretical terms, we would like to extend the term fragmentation, which Hindmarsh et al. used to describe observation of a discrete and limited fragment of the shared environment at any one time, by including to what extent the working space is shared. For example, the remote room in a videoconferencing setting is a fragment that can be observed, but not entered or interacted with; desktop CVEs support a fragment of the working environment, which can be shared but not physically entered; whereas IPTs support an interactive fragment that can be physically entered by all. We argue that within the already fragmented working environment created by a desktop CVE, further fragmentation of the shared fragment is caused by the low field of view and unnatural control of movement and gaze, however, we postulate that the same is not true for IPT based CVEs. Recent studies with linked IPTs have reported not noticing the changes of human behaviour induced by fragmentation, notated by Hindmarsh et al., although the experiments were not designed to isolate this phenomena [RWO*04, HSS*05].

The aim of this work is to contribute to the understanding of why IPTs seem better at supporting distance team work, through testing their impact on fragmentation and the effect on mutual orientation towards objects against that of desktop systems in an adaptation of the well known Hindmarsh et al. study. This initial study does this by measuring task performance and objective observation across a set of similar tasks that differ in spatial extent to compare collaboration in a CVE through desktops with that through IPTs.

This paper adopts the classic structure of following this introduction in section 1, by defining the experimentation in section 2, presenting results in section 3 and drawing conclusions in section 4.

2. Experimentation

The experiment studies the use of Immersive CVE technology in the collaborative task of organising furniture within a room. It is based on that of Hindmarsh et al. [HFH*00] and extends it by replacing desktop displays by IPTs and comparing task performance. This short paper only reports on the quantitative measure of task performance.

2.1. Environment

The environment has been modelled to closely resemble that of Hindmarsh's experiment. A large room is cluttered with a collection of chairs, some of which are visually distinct, and other more distinctive furniture, such as a standing lamp, computer-desk, television and HiFi system. Most of the furniture could be uniquely identified verbally but it was suspected that reference to it within the context of the scene and particularly through referential gestures would improve the performance of communicating its identity. Each object can be moved within the environment through direct manipulation by a single user. Figure 1 shows the layout of the environment and gives an indication of the active area during each task. Here, the numbers identify the object used in each of the five sub-tasks described below.

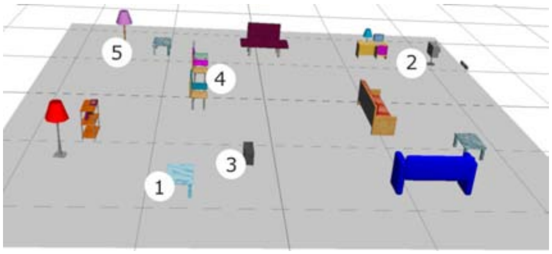

Figure 1: *Layout of the environment. The numbers indicate the area of attention within each task.*

2.2. Task

Like the earlier experiment [HFH*00], organising the furniture within the room is divided into a number of sub tasks. From previous experiments we suspect to see an improvement in task performance when using the IPTs. As we have simplified the collaboration by only allowing one user to move each object, we suspect the remaining major factors to effect task performance to be: effect of fragmentation on mutual orientation towards objects; naturalness of navigational movement and naturalness of interaction. We have therefore designed a set of sub-tasks distinct in scale of navigation and complexity of interaction, and raked them accordingly, from easy to hard:

1. Look to the marble table (referencing only)
2. Move the television to near a desk (some navigation and simple object manipulation)

Figure 2: *The sub task "Move the television" in the IPT-Desktop configuration: planning the task seen from the desktop (left), placing the TV as seen from the IPT (middle) and the TV in the final position as seen from the desktop (right).*

3. Move the HiFi system onto a table (some navigation and precise object manipulation)
4. Rearrange the chairs (medium length navigation and a series of object manipulations)
5. Move the stand lamp across the room (long navigation with simple object manipulation)

If we find that the level of improvement is proportional to scale of movement or complexity of interaction then we need to further isolate the effect of fragmentation on mutual orientation. However, if we find that the improvement is inversely proportional to both, we can deduce that the probable major factor in improved performance is the effect of reduced fragmentation on mutual orientation towards objects.

The collaborators were given leader and helper roles. At the start of the experiment, the leader enters the environment in the centre and moves to one side and then the helper enters in the centre.

The sub-tasks can be compared in terms of typical methodology as follows:

Bring attention to the marble table: The leader references the marble table, usually through verbally describing its appearance and position while turning and pointing to it. The helper looks at the leader to see where he is looking at or pointing to and turns in the indicated direction to locate the table.

Move the television near to a desk: the leader references the television and asks the helper to move it to the new location close to a desk. The helper then moves to the television, picks it up, carries it to the destination and places it on the ground. The leader decides if the object has been moved to the right place and directs the helper accordingly. See Figure 2 for an illustration.

Move the HiFi system to the marble table: This task is almost identical to the former, but differs in the intricacy of interaction as the helper must place the HiFi on top of the table as opposed to next to it.

Rearrange the chairs: The leader brings the attention of the helper to the set of four chairs towards the centre of the room and asks for all four to be shuffled. The helper moves to them and obliges. This sub-task requires much more varied movement from the helper, but within a relatively localised space that the leader can observe without turning. The leader calls a halt when he is satisfied with the result.

Move the standing lamp to across the room: This differs from moving the television and HiFi in the distance that the object must be moved and therefore the amount of room space the leader must observe during the operation.

Referencing is supported through capture and faithful remote reproduction of combinations of natural communication, such as turning to, pointing, touching, reaching for and picking up, and describing. This is represented through a remote avatar that follows the movements of the user within the same spatial context. Thus, if someone glances to a table within their working environment, the avatar will glance at an identical table, in just the same spatial context in the collaborator,s working environment. In this way, the object and meaning of a natural non-verbal gesture, be it conscious or unconscious, is communicated faithfully.

2.3. Display configurations

Two display types have been used: desktop and IPT. These were paired into three configurations of desktop-desktop, IPT-desktop and IPT-IPT. The IPT trials were carried out between displays at the Universities of Salford and Reading in UK. All other trials were carried out at Salford. The two desktop interfaces comprised an eighteen inch monitor, a 6 degree-of-freedom spacemouse and keyboard for navigation, microphone and speaker. Both IPTs comprised a three 3x3 feet wall and floor display, motion tracking of head and primary hand, wand with joystick navigation control, microphone and speaker. The Salford IPT used magnetic tracking technology, while that at Reading used a combination of ultrasonic and gyroscopic. In the IPT tests, the leading user was always based at Salford.

2.4. Embodiment

In all cases, the remote user was embodied by a jointed avatar. This character was modelled using realistic dimensions for limbs and a 3D scan for the face. Image capture and texture mapping was used for face, other skin, hair and cloths. Movement of the avatar is controlled through three points: head and both hands. These were tracked within the IPTs, and controlled through the spacemouse on the desktop systems. However, one of the hands was not connected to input for this experiment to simplify comparison between desktop and IPT. Torso position and orientation, as well as articulated arm movement were then improvised from two tracking points. Inverse kinematics were used to improvise arm articulation. The facial expression was static. The avatar could nod and shake its head and represent gaze, however, any turn of head in excess of thirty degrees results in the turning of the avatar's body to follow the head. Audio communication augmented the visual embodiment, but was not spatially tied to it.

2.5. Platform

The Immersive Collaborative Environment (ICE) [WRO04] was used as a test platform. We originally intended to use a more widely adopted test platform, but found that those available were unable to render the populated and cluttered environment at frame rate above the level of human perception on the available graphics computers.

2.6. Computers

The desktop PC systems were single processor machines, whereas the IPTs were run from SGI Onyx2 multipipe multiprocessor computers. Reading ran four walls from two pipes and used four processors, whereas Salford ran each wall from a separate pipe and used twelve processors.

2.7. Network Conditions

Typical ICMP ping tests between the desktops and between desktop and IPT were one millisecond. Similar tests between the two IPTs yielded around 17ms, as they were hundreds of miles apart and connected over the Internet. Network conditions were typical at the time of the tests.

2.8. Subjects

Sixteen voluntary test subjects have been taken from MSc students and their friends. All had prior experience of computers and around half have knowledge of the principles of VR and some prior experience of IPTs. The gender distribution was roughly equal. At the time of writing, eighteen trials have been undertaken, but only three of these within the two IPTs. People were shown how to use the system and given time to become accustomed to it before data was collected on their activities. This typically took around five minutes.

2.9. Measurement

Although we have recorded conversations and measured user experience through a qualitative questionnaire, this data has not yet been fully analysed and we restrict this paper to the quantitative measurement of task performance in terms of the time taken to complete each task.

3. Results

The field of view in the IPTs was usually above 120 degrees as the users had been trained to orient themselves towards the front screen most of the time, leaving a full side screen to each side and hiding the entrance behind them. Changes in field of view also mapped within the level of human perception, to changes in position, orientation and gaze. Eye saccades control the eye's fix point while bringing the head to a final position in a change of gaze and this is naturally supported in an IPT. Changing viewpoint on a desktop does not have this natural control system and is therefore more cumbersome. Furthermore, the desktop field of view was less than half that of the IPT. Finally, the user is actually within the shared environment when using an IPT as opposed to looking into it when using a desktop. Combining these factors we can strongly argue that fragmentation would have been grater when using the desktop displays.

Table 1 shows the measurements of time taken in seconds for each user pair to complete a subtask when interacting through the various display combinations. One can recognise a large deviation in the values where desktop displays were involved. When observing the collaborating users how they interacted through the various interfaces, we could detect similar divergent behaviours.

During this trial and others [RWO*04,HSS*05] it was observed that glancing around the environment in the IPT is very natural, provided the user's body is facing mostly towards the central wall in a three wall configuration. In comparison, viewpoint changes are cumbersome on the desktop and only occur as a conscious and deliberate action. Navigation within the immersive environment produces smoother and more efficient trajectories. On desktop displays, where the viewpoint is tied to the avatar, the user keeps stopping during navigation to look around and adjust the trajectory. When using the IPTs, people can be seen to follow the gestures and gaze of others with their own gaze naturally.

Figure 3 shows a graph of the results in Teable 1. The graph clearly shows a consistent impact of display configuration on task performance. In all five tasks, exclusive use of IPTs outperforms an IPT-desktop pair which, in turn, outperforms exclusive use of desktops.

Table 1: *Measurements of task performance within the distinct display configurations.*

Task		Time taken in second of user pairs								Average
Desktop - Desktop										
1	Look to the marble table	136	33	43	47	61	18	21		51
2	Move the television near a desk	90	61	60	93	49	55	72		69
3	Move the HiFi system to the marble table	80	72	70	107	42	60	37		67
4	Rearrange the chairs	95	112	102	130	145	120	85		113
5	Move the stand lamp across the room	117	70	78	73	38	79	69		75
IPT - Desktop										
1	Look to the marble table	63	41	22	43	20	50	16	24	35
2	Move the television near a desk	38	32	112	82	25	46	47	63	56
3	Move the HiFi system to the marble table	71	40	85	98	58	42	52	34	60
4	Rearrange the chairs	186	160	80	120	113	76	98	75	114
5	Move the stand lamp across the room	50	52	88	60	70	31	67	56	59
IPT - IPT										
1	Look to the marble table	18	25	20						21
2	Move the television near a desk	26	50	80						52
3	Move the HiFi system to the marble table	32	45	70						49
4	Rearrange the chairs	53	60	90						68
5	Move the stand lamp across the room	21	30	40						30

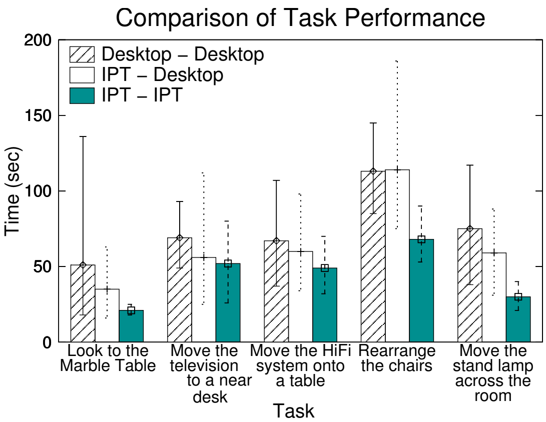

Figure 3: *Task performance compared over display device configurations.*

Table 2: *Percentage increase in task performance in using only IPTs over using only desktop displays.*

Percentage	Task
60	Look to the marble table
59	Move the television to a near desk
40	Move the HiFi system onto a table
27	Rearrange the chairs
24	Move the stand lamp across the room

The differences between the improvement across the tasks should tell us the impact the display is having on interaction and navigation. This is more clearly seen in Table 2 that shows the improvement in task performance as a percentage. One can see that the improvement reduces with scale of navigation and object manipulation, suggesting that better mutual orientation and thus a reduction in fragmentation that is leading to the task performance.

4. Discussion

As we might expect from previous trials, these results show an increase in collaborative task performance within a shared 3D space arising from exclusive use of IPTs. We have previously found that the natural use of the body in IPTs to reference and interact with objects increases both task performance and subjective impression of collaboration [RWO*04]. In order to isolate the effect of fragmentation on the efficiency of mutual orientation to an object of interest, we varied complexity of interaction and navigation. We found that task performance was always better in IPTs, but that the advantage diminished with an increase in navigation or complexity of object placement. We therefore deuce that the major factor in improved task perfromance is the effect of reduced fragmentation on efficiency of mutual orientation toward an object of common interest. At first glance, our results appear to conflict with our previous results that showed improvements in collaborative tasks focused on interaction with objects. However, in a parallel study, about to be presented at this year's Presence Workshop [ORW05], we have shown that object placement, is often easier through a desktop display, however, it is easier to see how someone else is interacting with the object if they are doing so through an IPT. In the study presented in this paper, we have removed the ability for collaborative manipulation of an object and so removed the need to accurately see how someone is manipulating it.

5. Conclusion

Problems with fragmentation and associated difficulties in mutual orientation to shared objects of interest have plagued all telecommunication systems to date. Observations of user behaviour within linked IPTs, in this and previous trials, have not found the same problems, but have found that most people use gaze and gesture in a natural manner. This study has added weight to these observations by showing a measured improvement in collaborative task performance, in a task reliant on the efficient referencing of objects, that can not be explained by improvements in naturalness of object manipulation or navigation. Together, all this suggests that IPTs have reduced the problem of fragmentation through placing people within the shared space and allowing natural use, capture and remote representation of gaze and gesture. We have previously demonstrated that it is easier to see how someone is working with a common object when they do so through an IPT. We have now shown that it is easier to identify which object they are referring to in a complex scene. Putting this together, strengthens our belief that the combination of wide field of view and the tracking of head and hand, reinforced by a strong sense of embodiment in the space, are characteristics of IPT that bring us closer to reproducing a face-to-face meeting than any other technology to date.

5.1. Future Work

At the time of writing we have not tested a sufficient set of people to provide conclusive results, however, these initial results are promising. A greater scale of test subjects is required to prove true statistical significance and we hope to have achieved this by the time of paper presented.

6. Acknowledgements

The authors wish to thank the UK research councils for funding the research and in particular HEFCE who part funded the infrastructure. We also wish to thank Vassil Alexandrov and his team at the University of Reading in UK for participation in the trial and Anthony Steed of UCL in UK, Dieter Kranzlmüller of Johannes Kepler University Linz in Austria, and their teams for assistance in the collaborative trials that led to this one.

References

[Art96] ARTHUR K.: Effects of field of view on task performance with head-mounted displays. In *Human Factors in Computing Systems* (1996), ACM, pp. 29–30.

[HFH*00] HINDMARSH J., FRASER M., HEATH C., BENFORD S., GREENHALGH C.: Object-focused interaction in collaborative virtual environments. *ACM Transactions on Computer-Human Interaction (ToCHI) 7*, 4 (2000), 477–509.

[HFHB] HINDMARSH J., FRASER M., HEATH C., BENFORD S.: Virtually missing the point: Configuring CVEs for object-focused interaction. In *Collaborative Virtual Environments*, Snowdon, D. and Churchill, E. F. and Munro, A. J. (Eds.), Springer, (2001), pp. 115–139.

[HSS*05] HELDAL I., STEED A., SPANTE M., SCHROEDER R., BENGTSSON S., PARTANAN M.: Successes and failures in co-present situations. *Forthcoming in Presence: Teleoperators and Virtual Environments 14*, 5 (2005).

[Kje01] KJELDSKOV J.: Interaction: Full and partial immersive virtual reality displays. In *IRIS24* (2001), pp. 587–600.

[LV02] LAPOINTE J.-F., VINSON N.: Effects of joystick mapping and field-of-view on human performance in virtual walkthroughs. In *The 1st International Symposium on 3D Data Processing Visualization and Transmission* (2002), pp. 490–493.

[ORW05] OTTO O., ROBERTS D., WOLFF R.: A study of influential factors on effective closely-coupled collaboration based on single user perceptions. To be presented at *The 8th Annual International Workshop on Presence*, September 21-23, London, 2005.

[PWBI98] POUPYREV I., WEGHORST S., BILLINGHURST M., ICHIKAWA T.: Egocentric object manipulation in virtual environments: empirical evaluation of interaction techniques. *Computer Graphics Forum 17*, 3 (1998), 41–52.

[RWO*04] ROBERTS D., WOLFF R., OTTO O., KRANZLMÜLLER D., ANTHES C., STEED A.: Supporting social human communication between distributed walk-in displays. In *ACM Symposium on Virtual Reality Software and Technology (VRST'04)* (2004), pp. 81–88.

[SSA*01] SCHROEDER R., STEED A., AXELSSON A., HELDAL I., ABELIN A., WIDESTROEM J., NILSSON A., SLATER M.: Collaborating in networked immersive spaces: as good as being there together? *Computers and Graphics 25*, 5 (2001), 781–788.

[WRO04] WOLFF R., ROBERTS D. J., OTTO O.: Collaboration around shared objects in immersive virtual environments. In *The 8th IEEE International Symposium on Distributed Simulation and Real-Time Applications (DS-RT'04)* (2004), pp. 206–209.