Shared 3D Workspaces

Joanna McGrenere
Kellogg Booth

July 1996

The University of British Columbia

1 Introduction

1.1 Background

In the last decade there has been an increasing demand for collaborative computer usage. People are requesting tools and technology that will allow multiple users to work together independent of their location; users can be in the same room, on the same floor, in the same building on a different floor, or even in a different country. The nature of the computer supported collaboration is varied. It ranges from people wanting to have the sense that they share personal space, such as an office, to wanting to work simultaneously on the same entity, such as a document or visual presentation. Even if technology is capable of such tasks there are significant human-computer interaction issues that arise. Protocols or metaphors for shared space need to be determined.

The need for such protocols or metaphors is perhaps best illustrated through the use of examples. Consider two people working together on the same document. Protocols for interaction have to be defined so that one user knows what part of the document is being edited by the second user in order that overwriting one another's work doesn't take place inadvertently. The computer needs to convey to each user what the other user is doing in a meaningful yet unobtrusive way. Another example is two people located in different buildings who are working together on a project. In order to facilitate both spontaneous as well as planned face-to-face communication, their offices may be linked with both audio and video connections. Interaction protocols are needed so that these links can be used effectively for collaboration while maintaining the privacy of the individuals.

The examples above illustrate collaborative computer usage where the shared space is 2-dimensional. The CSCW (Computer Supported Cooperative Work) literature includes considerable research that has been done in the area of 2D shared spaces and metaphors. Although research continues in this area, there is a new topic that is beginning to capture some research dollars. This is the area of 3D shared spaces. A logical approach to researching this topic is to extend our knowledge of the known to the unknown. Some questions that naturally arise are: *Can a 2D metaphor be used in 3D?* and *How is interaction different in 3D than in 2D?*

Metaphors for 2D interaction are not the only basis for 3D shared interaction. There has been considerable research in the area of 3D single-user interaction that is touched upon in the HCI literature and covered thoroughly in the virtual reality literature. Thus the next question that emerges is: *Can metaphors for single-user 3D interaction be extended to shared 3D interaction?*

It is clear that research done in the area of 2D shared spaces and 3D single-user spaces could provide important insights into the domain of 3D shared spaces. These areas should be explored before delving directly into research for 3D shared workspaces.

This report provides this exploration by way of a literature survey. Some key journal articles from the above two domains as well as a single article that documents shared 3D spaces are summarized. A discussion based on the articles is presented and areas that require further research are identified.

1.2 Overview of the Paper

Section 2 defines some of the terminology that is relevant to shared 3D workspaces. Section 3 provides the summaries of the articles that form this literature survey. Section 4 presents a discussion of the pertinent issues that are identified in the literature and notes areas requiring further research. Lastly, Section 5 provides a brief summary of this report. Note that if the reader is familiar with this area of research, the terminology and the article summaries can be skipped and the discussion section can be read directly.

2 Terminology

groupware: Groupware is software that explicitly supports group work. It is a technically-oriented label meant to differentiate "group-oriented" products explicitly designed to assist groups of people working together from "single-user" products that help people pursue their isolated tasks [4].

CSCW: Computer Supported Cooperative Work (CSCW) is the scientific discipline that motivates and validates groupware design. It is the study and theory of how people work together, and how the computer and related technologies can or do affect group behaviour [4].

media-space: A media space is a system that uses integrated video, audio, and computers to allow individuals and groups to work together despite being distributed spatially and temporally [18].

telepresence: Telepresence is the use of technology to establish a sense of shared presence or shared space among geographically separated members of a group [2].

shared-person space: Shared-person space in telepresence is the collective sense of copresence between or among group participants [2].

shared-task space: Shared task space is a copresence in the domain of the task being undertaken [2]. This can also be referred to as *tele-data* [4].

stereoscopic: Stereoscopic means that a presentation has different images for each eye and these different images are dependent on the position of the respective eye [14]. This term is often shortened to stereo.

head-coupled display: A head-coupled display is a display in which the calculation of the viewing transformations is based on the position of the user's head [14]. This is also called a head-tracked display.

head-mounted display: A head-mounted display presents images to one or both eyes through the use of small displays located on or near the head with appropriate lenses so that the images are seen as if viewing the world through glasses [12].

3 Paper Summaries

The first two papers provide a general introduction to telepresence as well as the task and person spaces. Desirable goals or features for these communication mediums are covered.

3.1 Beyond Being There [1]

This paper effectively argues the extreme position that we should not necessarily try to emulate face-to-face communication, or *being there*, when we design computer supported communication. We should instead design communication systems that are *beyond being there*, or better than being there. Stated in other words, the goal should be to design systems where people at a distance are not at a disadvantage to those who are present. For people at a distance not to be at a disadvantage, local people must use the system as well. The only way that local people will choose the system is if it offers more than meeting face to face. The latter is considered by the paper to be the litmus test of a system. The features that a new communication medium could take advantage of in order to meet this challenge are: the ability to support asynchronous communication; the ability to support anonymous communication; and the ability to automatically archive communication. One such form of communication that meets this litmus test is e-mail. The paper suggests the possibility of others.

3.2 Telepresence: Integrating Shared Task and Person Spaces [2]

This paper looks at both the task space and the person space and discusses the need for integrating these spaces and the issues that arise when they are integrated. Their seamless integration is one of the most important attributes of any telepresence system. In face-to-face interaction, these spaces are naturally integrated and the goal is to make the telepresence system as natural as possible.

The paper gives a number of examples covering both spaces as well as the integration of the spaces. A typical example of the person space is video conferencing. A major problem with many systems is the inability to establish eye contact among participants, a powerful and natural interaction cue. This can easily be corrected using teleprompting. Another small but effective adjustment is setting up the video monitor in portrait instead of landscape orientation which affords access to more body language. The Hydra system is introduced. It is a system in which each remote participant is represented by a separate video surrogate which has a separate camera, monitor, speaker, and microphone. The monitors are placed in the same order as the participants as though they were sitting in the same room. Thus person space is preserved. It is easier to maintain awareness of who is “visually attending to whom” and gestures such as head turning.

The same video channel that is used for the shared-person space can be used for the shared-task space. This channel can even be augmented by drawing on the video as is often done by sportscasters. A whiteboard type of implement can also be used. Finally, for the integration of the spaces, the example of *Shared ARK* is given. One of the most interesting things found with the experiments run on this system is that when visual attention was directed at the computer screen, it was found that the speech and non-speech audio established a shared space which was more effective than the highest fidelity video display. It was effective because the overhead in switching contexts was the same as in everyday life. Despite the title, however, the integration of the two spaces did not get enough coverage in the paper.

The next three papers cover different types of shared task applications. Tivoli permits collaboration where the participants are co-present. GroupSketch and SASSE represent task spaces such that the collaborators are remotely located. Design and interaction issues for these applications are discussed.

3.3 Tivoli: An Electronic Whiteboard for Informal Workgroup Meetings [3]

Tivoli is an electronic whiteboard application designed to support informal workgroup meetings. It runs on a Xerox Liveboard which is a large screen, pen-based display and it is targeted to support relatively small meeting sizes of up to eight participants. The designers' goal was to provide unselfconscious and fluid interaction between the users and the board; the board should not draw the participants' attention away from their interaction with each other and the board should enable the unhindered expression of ideas. Another goal was for the board to initially behave in a simple manner. This allows first-time users immediate use of the board and it also allows users to build from their current work practices involving whiteboards.

The paper documents the design features and issues that were considered in the development of *Tivoli*. It was found that a pen is more appropriate than a mouse because a pen enables both pointing and writing. Another issue was the placement of tools, such as buttons and menus as well as messages, on the board given the very large interactive surface and the proximity with which users would be to the board. A very important design issue was how to enable different users to use the board. It was decided that the board would support three different pens and that different modes were necessary. The two modes were pen state and system state. Something that belongs to a given pen's state, such as the selection of objects by that pen, could not be operated on by a different pen. Another example is that a pen could not erase the artifacts of a different pen. This seems to conflict with the designers' goal of making the board act as a regular whiteboard. In the latter case, a user can easily erase or modify objects drawn by a different user.

It is interesting that in the end the designers determined from user feedback that their interface design had in fact erred in favour of increased functionality over intuitiveness. They had lost the ease of use that a normal whiteboard affords. The designers recognized this as a major problem and intended to rectify it with the next release.

3.4 GroupSketch: A Multi-User Sketchpad for Geographically-Distributed Small Groups [4]

This paper focuses on a groupware system called *GroupSketch* which is a multi-user sketchpad supporting remote design activities by small groups. It allows users to list, draw, and gesture simultaneously in a communal work surface, supporting interactions similar to those occurring in the face-to-face process.

GroupSketch was designed based on six criteria formulated by Tang; its success is largely attributed to adhering to these criteria. Tang's criteria were derived from observations that he made during his ethnographic study of eight short small-team design sessions. The criteria are as follows:

1. Provide ways of conveying and supporting gestural communication. Gestures should be clearly visible, and should maintain their relation with objects within the work surface and with voice communication.
2. Minimize the overhead encountered when storing information.
3. Convey the process of creating artifacts to express ideas.
4. Allow seamless intermixing of work surface actions and functions.
5. Enable all participants to share a common view of the work surface while providing simultaneous access and a sense of close proximity to it.
6. Facilitate the participants' natural abilities to coordinate their collaborations.

In brief, *GroupSketch* supports four participants. Each participant has a labelled cursor and a unique caricature displayed outside the border of the writing/drawing space. There are four action modes, namely pointing, drawing, listing, and erasing, and the cursor automatically changes form depending on the mode. The mode is indicated by the input. For example, to draw, the mouse is used with the left button depressed. All different forms of the cursor are extra-large in size; they are 64bit by 64bit instead of the regular 16bit by 16bit. This permits better visibility and coordination of participants' activities. There is no social protocol enforced for the interaction; the coordination is left entirely to the participants. There is a fully duplex audio channel enabled. The system provides instantaneous shared views of the display.

Based on observation, *GroupSketch* proved to be well liked and easy to use. The aspect most disliked was having to use the mouse to draw. This could easily be rectified by using a pen instead. Participants who were more computer literate expressed the desire for more functionality. With added functionality, however, ease of use is often lessened.

3.5 The User-centred Iterative Design Of Collaborative Writing Software [5]

This paper documents user-centred design based on behavioural research that was used for the development of the writing software *SASSE*. This software permits multiple users to write, edit, and review documents synchronously or asynchronously. For the purposes of this report, much of the research that is documented is not of interest. What is interesting, however, is the discussion on collaborator awareness.

Collaborator awareness is achieved through colour, views, audio, and instantaneous update. Each author is assigned a unique colour at document creation time. This assignment is stored with the document so that a given author has the same colour each time a particular document is edited. Users can determine where the other authors are working by colour-coded text selections and scroll bars. Collaborator awareness is further aided by views that provide information about the state or actions of collaborators. The *gestalt view* presents a condensed image of the entire document as well as all collaborators positions and text selections. The *observation view* allows users to “look over the shoulder” of a fellow collaborator to see exactly what they are seeing and doing. Non-speech audio cues provide information about collaborators’ actions such as scrolling and deleting. Lastly, having all participants’ workstations updated instantaneously when updates/changes/additions are made to the document enhances collaborator awareness.

The following two papers cover person space issues. They both extend the definition of collaboration and present methods to support this full form of collaboration.

3.6 Working Together, Virtually [6]

The majority of research in computer supported collaborative communication has been for the facilitation of meetings or what may be considered formal communication. Comparatively little has been done in the area of frequent and spontaneous informal communication. This paper introduces the concept of a virtual open office to address this deficiency. The goal of the virtual open office is to enable closeness and cohesion of co-workers engaged in joint work. It should enable communication of participants within the same building or at a greater distance.

One of the most important contributions of the paper is the eleven user requirements that need to be met in order to achieve a virtual open office. They are the ability: to implicitly establish a co-worker’s level of accessibility; to enforce reciprocity in information exchange; to explicitly set one’s level of accessibility; to trivially make and close verbal and visual contact; to have multi-way conversations; to support multi-media information exchange; to filter out unwanted noise; to discriminate among sounds in the virtual open office; and to obtain feedback on the communication environment. A virtual open office called *VOODOO* is introduced that meets the above requirements. Generally *VOODOO* appears to be a feasible solution to the problem of supporting informal communication. There are a couple of design features, however, that seem somewhat impractical or unnatural. For example, having a video image of everyone in the virtual office on a terminal would be problematic unless the virtual office included only a small handful of people. An example of an unnatural interaction is the need to click on all the participants’ images who are not supposed to overhear a conversation. Perhaps selecting participants while in a “whisper” mode would be a better solution.

3.7 Realizing a Video Environment: Europarc's Rave System [7]

This paper introduces the *Ravenscroft Audio Video Environment* (*RAVE*) which is a media space that is designed to enhance the working environment and promote collaboration. Similar to *VOODOO*, the designers of *RAVE* have significantly broadened the meaning of collaboration. Collaboration is not considered to be two or more people focused intensely on a single task but rather anything from spontaneous to highly planned communication or from disengaged to highly focused communication. There are *RAVE* functions, called buttons, that reflect the range of engagement in the levels of collaboration just mentioned. The five buttons are *background*, *sweep*, *glance*, *office share*, and *vphone*. The *background* button allows people to select a view from one of the public areas to display on their monitor. *Sweep* provides a way to maintain awareness of remote locations of the building by making approximately 1-second one-way connections to various nodes in the building. *Glance* makes single 3-second one-way connections to a selected node and allows more focused attention at particular colleagues. This is often used to see if a particular person is busy. *Vphone* is similar to a telephone call except that both two-way audio and video connections are established. *Office share* is similar to *vphone* except that the connections are meant to last significantly longer. It is supposed to provide the same effect as sharing one's office.

Clearly the main disadvantage of the above audio-video system is the threat to privacy that it poses. Enforcing symmetrical two-way connections so that seeing or hearing somebody implies being seen or heard oneself was rejected as a means of protecting privacy. It was decided that enforcing symmetry for the sake of privacy would undermine the ability to use video as a means to gain general awareness unobtrusively. The paper addresses four facets of privacy: the desire for *control* over who can see or hear us at a given time; the desire for *knowledge* of when somebody is in fact seeing or hearing us; the desire to know the *intention* behind the connection; and the desire to avoid connections being *intrusions* on our work. *Control* is handled by the architecture which allows a user to select the functions that will be available to other users where s(he) is the target of the functions. *Knowledge* is covered by auditory notifications. These notify a user when another user tries to make a connection to their node. *Intention* is revealed in the form of notification that occurs; different auditory cues are used for the different functions.

The next six papers cover interaction in 3D space. Virtual Reality is introduced. Single-user interaction and multi-user interaction are covered.

3.8 Facile 3D Direct Manipulation [8]

This paper documents an experimental 3D interface for object manipulation that achieves casual, direct, and natural 3D interaction. The interface, according to the author, attempts to isolate the complexity in the computer and not in the interface or in the user's mind.

The paper introduces a 3D pointing device called a roller mouse. It is a standard one-button mouse with two wheels on the front. These wheels move the cursor closer and farther from the camera. Both wheels perform the same function. The reason for having two is to accommodate both left and right-handed users. If an object obscures the cursor, then the object is rendered translucent. When the cursor "touches" an object, cross-hairs appear within the object. This is essentially the same as selecting the object. The cursor is normally in a cone shape with the tip of the cone indicating the direction that the cursor is moving. When the cursor is in a cone shape and an object is "touched," then the object can be translated and rotated simultaneously by click-dragging on it. A technique called tail dragging is used to control the rotation and the cursor controls the position directly. When an object is "touched" and the cursor is moved towards the center of the object, the cursor turns into a jack shape. This signifies that the object can only be translated by click-dragging on it and moving the mouse and wheels. Experiments indicated that users found the 3D mouse to be a natural extension of a 2D mouse and were quite easily able to control the cursor, and even master the complex interplay between the mouse body, the button, and the wheels.

A method called *snap-to* is introduced as a manipulation technique. It uses an intuitive model of magnetic attraction to help users align objects in both position and orientation. As an object is moved towards another it is pulled away from the cursor and toward the attracting object. The effect is reinforced visually by a small red spring that compresses as the objects near one another. In addition, audio reinforcement is successfully used to accentuate the snap-to interaction and many other interactions in the interface.

This paper documents a first attempt at this form of direct 3D direct manipulation. Issues of object size and what happens if the cursor cannot fit into the object, are not addressed.

3.9 A Space Based Model for User Interaction in Shared Synthetic Environment [9] *and* Integrated CSCW Tools Within a Shared 3D Virtual Environment [10]

These papers present a model in which 3D space is used to provide an interface to applications and resources. This 3D space is a virtual reality environment in which the different participants are represented as stylized 3D icons. The metaphor for interaction is based on the concept of presence or proximity. Proximity is modeled in the virtual environment with a geometric volume of the immediate surroundings called the aura. Proximity is then used to establish communication channels, to establish presence at meetings, and to provide interaction. The goal for this system is to use direct, real world metaphors so that interaction and communication are made as natural as possible which results in minimal cognitive load being placed on the users.

For example, when the aura, or geometric volume, of a user intersects with that of another user then the communication channels between these two users is automatically opened. Aura is also extended as a property of services and tools. So if the aura of a user intersects the aura of a workspace, such as a whiteboard or conference room, then the user is able to use the services of the object. These services include the ability to write on a whiteboard, to pass out documents at a conference table, to speak at a podium, etc. Both the opening of communication channels as well as the availability of services occurs transparently to the user when the auras intersect.

The system *Distributed Interactive Virtual Environment* (*DIVE*) was built to accommodate the above 3Dspace. At the time the paper was written, however, sound and video had not yet been implemented. If a significant number of users participate at a meeting, establishing all the audio channels among the participants will be a significant challenge. Video images of the users are also crucial. The stylized icons, which are currently used, are somewhat humourous looking and would not be conducive to having a serious meeting.

3.10 Nature and Origins of Virtual Environments: A Bibliographic Essay [11]

This paper provides a broad overview of the many facets of virtual environments. As it is a summary of the topic, I can hardly do the paper justice by summarizing its summary. Instead I will highlight an area that is covered in the paper that pertains to the report at hand, namely, the three different levels of virtualization.

Virtualization is defined by the paper as “the process by which a human viewer interprets a patterned sensory impression to be an extended object in an environment other than that in which it physically exists.” The three levels of virtualization distinguished by the paper are: virtual space, virtual image, and virtual environment. Virtual space is created when a user perceives objects laid out in three-dimensions when in fact viewing a flat surface that presents perspective, shading, occlusion, and texture cues to the space. The second form of virtualization is the perception of a virtual image. It is defined as the perception of an object in depth in which accommodative, vergence and (optionally) stereoscopic disparity cues are present, though not necessarily consistent. Lastly, a virtual environment has the added information of observer-slaved motion parallax, depth-of-focus variation, and wide field-of-view without a prominent frame. The difference between these levels, I believe, is the immersiveness of the virtual reality.

The paper notes that virtualization is essentially a communication medium and as such is intrinsically applicable to practically anything from education to scientific visualization. Because the goal is effective communication, however, it is important that the virtual system selected for a given task be appropriate for the task. Despite the allure of creating an alternate reality, one need not aspire to creating a fully implemented virtual environment: a virtual space or a virtual image might even be superior.

3.11 High Resolution Virtual Reality [12]

This paper explores issues related to implementing head-tracked stereo displays. Four issues arise when moving from an image on a flat screen to pairs of images on the viewer's retinas. The first is the need for predictive head-tracking in order to reduce lags. This involves a forward prediction of where the user's head is likely to be when the rendering and display of the next frame is completed. The second issue is the dynamic optical location of the viewer's eyepoints. Because the real eyepoint does not lie at the center of rotation of the eye, the exact location of the eyepoint changes slightly depending on the direction of the user's gaze. The error due to uncertainty in eye nodal point location can be minimized by guessing where the user's eyes will be. The paper makes the guess that if a 3D mouse is used, the gaze will be in the direction of its "hot spot." The third issue is determining physically accurate stereo perspective viewing matrices. These matrices determine the relative position, orientation, and scale of the superimposition of the virtual world onto the physical world. The physical configuration of the stereo display device and the sensed real-time location of the user's eyes contribute to these matrices as well. The final issue addressed is the need to correct for the refractive and curvature distortions of glass CRTs.

Experimental results show that if these four issues are addressed then it is possible to achieve registration accuracy of less than a centimetre when the computer-generated worlds are superimposed onto the physical world. Not only does this allow users to use their normal binocular vision to accurately judge distances to virtual objects, it also allows virtual and physical objects to be intermixed.

3.12 Fish Tank Virtual Reality [13] *and* Evaluating 3D Task Performance for Fish Tank Virtual Worlds [14]

The definition of fish tank virtual reality is "the use of a standard graphics workstation to achieve real-time display of 3D scenes using stereopsis and dynamic head-coupled perspective" [14]. Both of these papers explore the advantages of fish tank virtual reality (VR) over immersion virtual reality. They also document experiments that were run to test the relative importance of head coupling vs. stereo display.

The first advantage of fish tank VR is resolution. Immersion VR employs a head-mounted stereo display. Because the screens in such a display are placed very close to the eye, the resolution is significantly worse than that of a high resolution monitor. Fish tank VR is also better able to simulate the effect of depth of field, has better stability in the presence of eye movements, and lastly, is more easily integrated into the everyday workspace.

A number of findings were observed from the first two experiments. Head coupling had a greater impact on performance than stereo and users consistently preferred head coupling without stereo over stereo without head coupling. When both factors were used, the result was an order of magnitude improvement in task performance compared to standard display techniques. In the last experiment it was found that lag is more important than frame rate in determining user performance and, in fact, frame rate itself is probably not important except for the lag it produces.

3.13 Decoupled Simulation in Virtual Reality with the MR Toolkit [15]

This paper documents the *Decoupled Simulation Model* (*DSM*) that can be used for creating successful VR applications and a software system, called the *Minimal Reality* (*MR*) *Toolkit*, that embodies this model. The objective of the model is to simplify the development of VR applications by providing standard facilities required by many VR applications.

DSM is based on nine requirements. The first five requirements are related to the interactive performance of VR applications and the last four are related to issues of creating software for virtual environments:

1. VR applications must generate smoothly animated stereoscopic images for head-mounted displays to maintain the key VR illusion of immersion. The application must provide a visual-update rate of at least 10 updates per second independent, if possible, of the application-update rate which can often take longer.
2. VR applications must have lags of under 100 milliseconds in order to be interactive.
3. VR applications must provide support for distributing an application over several processors.
4. The toolkit should provide an efficient data communications mechanism and should hide as many of the communications details from the programmer.
5. Performance evaluation of VR applications is needed
6. Applications should be portable from one site to another.
7. The toolkit should provide support for a wide range of input and output devices.
8. Applications should be independent from room geometry and device configurations.
9. A flexible development environment for VR applications is needed.

The *MR Toolkit* which is based on *DSM* has facilities which include support for distributed computing, head-mounted displays, room geometry management, performance monitoring, hand input devices, audio feedback, and data sharing.

Here a transition is made away from the topics of shared workspaces and graphical displays. These last two references cover selected topics in the area of visual attention.

3.14 Abrupt Visual Onsets and Selective Attention: Evidence From Visual Search [16]

This paper documents three experiments that show that isolated abrupt onsets are rapidly detected in visual search. Such onsets refer to single targets or objects that appear abruptly on a display. The paper also presents an attentional capture model that satisfactorily accounts for the data collected in the experiments.

The general design for each trial in Experiment 1 is that the subject has to locate a given target. No-onset items appear on a display “gradually” in that they are camouflaged and are then uncovered. At the same time that these items become fully uncovered, a single onset item appears. The target is either one of the no-onset items, the onset item, or not on the display at all. The results show that on every trial of visual search, attention is rapidly focused first on the abrupt onset location and then all other locations are scanned serially in a random order until a target is found or the search is complete. In Experiment 2 the possibility that subjects could more easily process the onset items than the no-onset items due to some physical difference between the two was tested and rejected. In this experiment, when the subject was told in advance where their attention should be focused, there was no difference in detection latency between the onset and no-onset items. In the third and last experiment, it was established that the no-onset procedure that was used in the previous two experiments was effective independent of whether the camouflage that “hides” the no-onset items is removed quickly or gradually.

The abrupt-capture model to which the data from the three experiments is fit is expressed as follows: $RT = A + kT + \delta N$ where RT is the response time, A is a random variable reflecting the time for all mental operations not accounted for by the other terms in the equation, k is the number of comparisons required on a trial, T is a random variable reflecting the time required to complete one comparison, δ is an indicator variable that equals 1 if the target is absent and 0 otherwise, and N is a random variable corresponding to the extra time required to deal with a negative trial (when the target doesn’t appear).

3.15 Some Primitive Mechanisms of Spatial Attention [17]

This paper documents the hypothesis that there is an early preattentive stage in vision where a small number of salient items in the visual field are indexed and thereby made readily accessible for a variety of visual tasks. A provisional model, called the *FINST* model, is developed to account for this spatial indexing. A number of studies were conducted to test the hypothesis. I will briefly highlight some of these studies.

The first set of studies covers multiple-object tracking. They show that it is possible to track about four randomly moving objects and to keep them distinct from visually identical distractors. It appears that it is the indexed items themselves that are tracked, rather than some contiguous region that contains them. The studies show that the indexed items may be tracked serially or may be tracked in parallel. What is definite, however, is that tracking is not being done by a process of scanning attention from one object to another in the total object space of both indexed and non-indexed items.

The cueing studies show that up to five items can be precued from among a larger set of similar items and that the cued items are treated by the visual system as though they were the only ones present. Cueing can be accomplished by the abrupt onset of markers which mark the selected items. It is shown that the selected set is searched in the same way, in parallel or serially depending on the type of items, that they would have been searched if they were the only items displayed. The studies also show that if the precued items are searched serially, it is not accomplished by a scanning process because greater spatial dispersion does not lead to slower response times.

A third set of studies investigated a phenomenon referred to as “subitizing” which is the rapid and accurate enumeration of a set of less than five items under certain conditions. Because the items can be counted much more quickly than it would take to count items that required serial attention, it is believed that a small number of indexes are assigned to the items and that subitizing is accomplished by merely counting the number of active indexes.

4 DISCUSSION

The above summaries cover a relatively broad range of literature, mostly from the human-computer interaction domain. Clearly not all of the material covered in all these articles is relevant for the topic at hand. Here I attempt to draw out the pertinent information from the literature and discuss the various themes that emerge. Where I can, I try to show what role this information plays in the domain of 3D shared interaction and I try to identify areas that require further research.

As a starting point it is important to have some concrete examples of where 3D collaborative systems may be useful. A few examples can be derived from the area of medical visualization. Consider the scenario of two or more remotely located heart specialists looking at a 3D image of a patient's heart and collaboratively making a diagnosis. Supporting this form of collaboration reduces the time and expense of travel that would otherwise be incurred for the specialists to visit the site of the patient. It also enables spontaneous second opinions from remote specialists in emergency situations that would not otherwise be possible. Another medical example could be two or more co-located doctors performing surgery. The doctors look into the patient from their various angles, and in addition to seeing the patient, they see a virtual image overlay of the patient indicating where the incisions and other activities must take place. The benefit of such support is that guidance is provided and errors are thereby minimized. A third possible example for 3D collaborative support would be to extend telepresence to 3D. Instead of only having the impression that you are sitting across from a colleague and looking into their office as is the case with the current implementation of telepresence, the user would actually feel as though they were sitting in their colleague's office. This would give a heightened sense of being there in the colleague's presence.

The notion of "being there" has been explored [1] and is relevant to the discussion at hand. If we want remotely-located people not to be at a disadvantage to the people present when using a computer supported collaborative system, then the people present must choose to use the system over meeting face-to-face. In other words, the system must offer more than just face-to-face communication. Ideally, if the 3D collaborative system were to offer such features as asynchronous communication, anonymous communication, or automatic archival then there is a higher likelihood that people in the same location would also use the system. Consider the diagnosis example from the previous paragraph. If the system could support asynchronous diagnoses and could store a specialist's diagnosis and perhaps an animation that steps through the problem areas in the heart as the diagnosis is explained, then local specialists would probably be inclined to use the system. This system would allow the different specialists to make their individual diagnoses at a time of their own convenience and fully document the diagnosis for the other specialists to view at their leisure.

The above features that make a system more useful than actually meeting face to face are very important but are also somewhat premature at this stage. First it is necessary to understand synchronous collaborative 3D interaction. This understanding can be drawn from research done in 2D shared interaction and in 3D single-user interaction. In order to extrapolate 3D shared interaction issues from these two domains of research it is important that a clear separation of shared-person and shared-task spaces be understood. These two spaces generally demand different interaction styles and thus it is easier to deal with them separately. This is not to say that the shared-task and shared-person spaces cannot be combined into one integrated space. Even when combined, however, it is easier for analysis purposes to distinguish the interaction that is related to the task space from the interaction related to the person space. It is clearly possible to have an integrated space where the two component spaces are in different dimensions. It would be possible to have a 3D task space while keeping a 2D person space. For example, with the heart specialist making remote diagnoses, it might be the case that the specialists attend to the task of viewing and diagnosing a 3D image of a heart while at the same time they discuss their diagnosis via video hookup which provides a 2D person image. In this particular case the only extension made from what is already available is the extension of the task space from two dimensions to three. The 2D person space is already covered in the literature and will not be covered in this report. Similarly, it is possible to have a 2D task space with a 3D person space. This would be the case if two people were working in a 3D telepresence environment on a 2D document. Again, the 2D task space is covered in the literature and, therefore, the 3D person space need only be addressed.

The goal is to understand and uncover interaction issues pertaining to 3D person spaces and 3D task spaces. First I will look at how 2D shared interaction can be extended to 3D shared interaction. Looking first at person spaces, the two main features documented for 2D person spaces are the audio and video connections. An important issue that arises from these features is privacy. As will be discussed in the following paragraphs, each of these features/issues would be similar if not the same in 3D as they are in 2D.

There are a number of existing systems that support 2D shared person spaces. The *RAVE* [7] system created at Europarc and *VOODOO* [6] at the University of Toronto are two examples of systems that are documented in the literature. Both of these systems were created in order to address the lack of support available for informal or spontaneous communication. They support both formal and informal communication. In order to establish a shared presence among two or more people both *VOODOO* and *RAVE* use audio and video connections. The audio is obviously used to enable vocal communication. It can also be used, however, to provide feedback as is seen with the *RAVE* system. For example, auditory notification is given to a user when a second user attempts to make a connection to the first user. This type of nonspeech audio has a number of advantages over graphics, text or speech: sounds can be heard without requiring the kind of spatial attention that a written notification would; non-speech audio cues often seem less distracting and more efficient than speech or music; sounds can be acoustically shaped to reduce annoyance; and finally, caricatures of naturally-occurring sounds are a very intuitive way to present information (e.g. the sound of an opening or closing door) [7]. Having said all this, audio is clearly a feature that is not specific to a 2D space and therefore no further discussion relating audio to a 3D space is necessary.

Video, unlike audio, has a definite dimensionality associated with it. The video images in both *VOODOO* and *RAVE* are 2D or flat images. In *VOODOO* the images are of the user in their office. *RAVE* includes the latter images as well as images of public areas in the building. This functionality that video affords could be extended into 3 dimensions. Instead of simply seeing a flat image of a colleague, video could provide the illusion that the user is actually seated across from their colleague. And instead of seeing a flat image of a public space, video could support the view that the user is actually standing in the space. Augmenting video from 2D to 3D represents a move into the realm of virtual reality. Virtual reality will be covered later in this report.

The relative merits of audio and video for creating a shared presence is a well discussed topic [4, 7]. The consensus seems to be that allowing people to see one another does not add significantly to the process of collaboration. In other words, visual information has no significant effect on the dynamics of conversation. However, tasks that involve conflict, bargaining and negotiation are affected by visual communication [18]. This leads to an important question: given that the role of video in a 2D shared space is minimal, is there any advantage to having a 3D video component in the shared space? The cost of moving from 2D video to 3D video could be quite significant not only in terms of bandwidth but also in terms of the number and sophistication of the cameras involved. Incurring this cost only to achieve minimal gain could prove to be very cost ineffective. It could be the case, however, that although 2D video contributes minimally, 3D video, for whatever reason, actually makes a significant contribution to the feeling of a shared presence. Clearly, this is an area that requires further study.

While audio-video connections provide a sense of shared presence, they can also pose a serious threat to the privacy of the individuals using them. As mentioned above, both the *VOODOO* and the *RAVE* systems have video connections directly into the users' offices. This enables users to glance into their colleagues' offices to see if they are there or to establish a connection for communication. Although this functionality is extremely useful, it is clearly gained at the expense of privacy. *VOODOO* addresses the privacy issue in two ways. First it allows the user to explicitly set a level of accessibility so that the user decides when and when not to permit interruptions. It also enforces reciprocity in the video and audio channels. It does so under the premise that in an open office, viewing and listening is reciprocal. The developers for *RAVE* have taken a slightly different approach. They have kept one-way connections because of the advantages that they provide. For example, glances provide awareness without actually engaging or interrupting one's colleague. Notification that a colleague is glancing is given, however, in the form of an auditory cue as described previously. Similar to *VOODOO*, *RAVE* allows the user to explicitly set their accessibility. In *RAVE*, however, accessibility can be set differently for different colleagues or co-workers.

The issue of privacy would be the same or perhaps even accentuated if video was presented in 3D. The issues of wanting to be aware if you're being seen by others and setting one's level of accessibility do not change based on the dimensions of the video image. What may in fact be more intrusive, however, is the number of cameras required to provide 3D video. Instead of having one camera mounted on the top of a workstation and perhaps a second mounted on the office wall, multiple cameras would be required. These cameras would also have to be somewhat mobile in order to provide views from many different angles. I would speculate that having multiple roving cameras in one's office could be extremely intrusive.

Moving from person space to task space. Similar to the person space, there are a number of features and issues in the 2D task space that would be much the same if they were extended to a 3D task space. These include: the use of colour; supporting different views; having audio feedback; the limited contribution of video; and supporting synchronous as well as asynchronous interaction. I will discuss each of these in turn.

Colour is used effectively to promote collaborator awareness in the shared text editor *SASSE* [5]. Each of the authors maintains a unique colour for the life of the document, and thus all document updates can easily be associated with a particular author. In addition, colour makes it easy to discern where in a document the different authors are working. *SASSE* also provides the use of telepointers. I would assume that the telepointers for each of the authors take on these same unique colours. So in addition to being able to locate where a particular author is working, it is possible to distinguish which author is pointing something out in the document for the other authors to focus upon. This use of colour could be used equally effectively in 3D. For example, if three different heart specialists were dissecting a 3D image of a heart, collaborator awareness would be enhanced significantly if each of the specialists were given their own uniquely coloured instrument (pointer). The design of 3D pointers is a research challenge.

Another way to increase collaborator awareness is through the provision of different views. A *common* view is an important view that was missing in some of the first shared task space systems [4]. The importance of such a view is covered in the fifth item of Tang's six criteria [4]. A common view enables all the participating users to orient themselves in the same direction towards the work surface so that all the users see the same thing. This view can also be described as an *observation* view [5]. Another view, called a *gestalt* view, presents a condensed image of an entire document as well as all collaborators' positions and text selections in the document [5]. This view need not be specific to a document task; it is applicable to working on any task in which the participants can work in different locations within the shared workspace, such as a whiteboard. Having these views would perhaps be even more important in a 3D task space than in a 2D space. Given the added complexities associated with an extra dimension, it could be extremely difficult for users to orient themselves so that they are seeing the exact same view as one another. It would, therefore, be crucial that the system provide the common view. Having a *gestalt* type of overview could also be extremely beneficial in 3D. If specialists were dissecting a 3D image of a heart it would be beneficial if they each knew where the others were working. Required views, apart from the two mentioned, are probably task specific. Given a particular 3D task application, the necessary views would have to be determined.

Audio feedback and video play a similar role in the task space as they do in the person space. In the same way that non-speech audio cues can indicate that a colleague has glanced into your office, these cues can also be used to indicate that a collaborator is scrolling or deleting when working on a document [5]. Audio cueing could be equally beneficial in a 3D task space. For example, a cue could be given when a specialist is making an incision so that the collaborators are aware that a cut is being made. Again, the actual audio will not change when 2D applications are extended to 3D and so no further discussion on audio is warranted.

Video can be used to support the performance of a task [2]. This usually occurs when one user is actually performing a task and the collaborating users are watching and perhaps providing help. Here the video is the actual task space. It is more often the case, however, that video is used in addition to having a 2D task application such as an editor or whiteboard. So the video is really adding a sense of shared-person space to the task space. Thus we have the integration of the spaces. A couple findings have been documented: the movement of the cursor synchronized with a participant's voice provides the greatest sense of telepresence [4]; and when visual attention is directed at the computer screen, the speech and non-speech audio establish a shared space which is more effective than the highest fidelity video display [2]. Given the limited contribution of 2D video when it is used to support a task space, it is highly unlikely that 3D video would provide better support. If the 2D video is the actual task space, however, it is certainly possible that 3D video would provide a better sense of actually performing the task or visualizing the task being performed. Further research in this area would be required.

Another feature of 2D task space systems that would be useful in 3D systems is the support for synchronous and asynchronous interaction. For example, synchronous writing is essential during the stages of brainstorming and outlining whereas support for asynchronous work is particularly important in the stages of writing, editing and reviewing [5]. These two different types of support could be useful in 3D as well. Consider again the example of the heart specialists making a diagnosis. The system should clearly support both the simultaneous diagnosis by the specialists as well as independent diagnoses made by each specialist that could be reviewed by the others at another time.

Thus far the extension of features and issues that pertain to 2D person and task spaces into 3D shared space has been discussed. There is, however, a different extrapolation that could be made. This is the extension of 3D single-user interaction features and issues into 3D multi-user interaction. 3D single-user interaction is primarily documented in the virtual reality (VR) literature. The volume of literature that covers VR is immense and is continuing to grow at a fast pace. The references that I have extracted are some of the key VR references but are by no means exhaustive of the subject area.

Three different levels of virtualization have been classified based on the level of immersion they provide. For the discussion at hand, distinguishing two different levels will suffice. These are immersion virtual reality (IVR) [12] and fish tank virtual reality (FTVR) [13,14]. Recall that in IVR the user is made to feel as though (s)he is fully immersed in the virtual world. This is usually achieved through a head-mounted display which presents images to one or both eyes through the use of small displays located on or near the head. FTVR is a less immersive form of VR in which the real world can be intermixed with the virtual world. This can be achieved with a regular CRT monitor and a head-tracked or stereo display.

Because the user is fully immersed in IVR, the task space and person space are necessarily merged. Given that true immersion is fully attained through the proper implementation of force feedback and other required mechanisms, interaction in IVR should ideally be the same as it is in the real 3D world. In other words, users should not need to learn how to interact. A type of immersed world called a *Space Based Model* has been documented [9,10]. In this model interaction is enabled based on proximity and aura, where aura is represented by a geometric form that encompasses the 3D stylized icon of the user and proximity is established when auras intersect. So if the auras for two users intersect, then they are proximate and can interact. Similarly, if a user's aura intersects that of a tool, such as a whiteboard, then the user is able to use the whiteboard. The *Space Based Model* presents a reasonable attempt at achieving full immersion but it would be difficult to argue that it comes remotely close to modelling the real 3D world. There is much work yet to be done in simulating reality in a VR world. Until this is achieved, interaction in VR will not be the same as interaction in the 3D world.

FTVR, according to the literature, has only been used for single-user task applications. So unlike IVR, FTVR operates in an environment where task and person spaces are separated. It should be possible, however, to integrate an FTVR task space with a 2D person space by adding 2D video or audio. It would be difficult to integrate a 3D person space with the task space, however, because of the nature of the head-coupled display and the limitations of the size of the CRT. Assuming for the moment that FTVR is achieved through both a head-coupled and stereo display, there are a number of interaction issues that arise. First, if a user wants to adopt the view seen by another user, i.e. a common view, how will the head coupling and stereo vision adjust? Will it simply be disabled until the common view is deselected? Another issue arises from the fact that FTVR can intermix with the real 3D world. The example given at the outset of this discussion regarding surgeons performing surgery on a real patient but with a virtual overlay exemplifies this intermixing of the virtual and real worlds. If the participants are remotely located, then task-related objects in one user's real world would have to be displayed as virtual objects to the other user. The system would need to provide the ability to update a real object if the remote user makes an update to the representative virtual object. These issues will need to be resolved through further research. Another interesting area for further research would be determining what IVR interactions are valid in FTVR.

There are two technical aspects applicable to both IVR and FTVR that should be mentioned briefly. These are frame rates and lags. In order to maintain the key VR illusion of immersion, the application must provide a visual-update rate of at least 10 updates per second, independent of the application-update rate which is often longer [15]. I would suspect that a similar update rate would be required even for FTVR. Further, applications must have lags of under 100 milliseconds in order to be perceived as interactive [15]. Maintaining these two conditions will be even more critical for multi-user VR applications. In order for a user to react to the motions made by a second remote user, it is essential that the user be made immediately aware of the second user's actions. More importantly, a user may try to alter or navigate an outdated copy of the 3D world if frame rates and lags are too slow. This is analogous to the need in 2D task applications for instant update.

The allure of virtual reality, especially IVR, is extremely high to say the very least. Thus, the potential to use VR in systems for its appeal alone definitely exists. Giving in to this hype may in fact lead to poorer systems [11]. Designers should carefully assess the purpose of their systems and decide what type of environment best supports the given purpose. Although IVR has a stronger appeal, FTVR could be more appropriate for the task even if an intermixing of virtual and real objects is not required. Or it could be the case that having no virtual component would in fact be best for the task.

Another research area quite separate from the ones previously mentioned that has a significant role to play in 3D shared interaction is attention psychology. There has been a large amount of work done in the area of attention and I am far from able to do it justice here. I have included two references on attention to give it a representation, albeit a small one, in this survey. The discussion I pursue below should not by any means be seen as exhaustive on the topic of attention.

It has been shown empirically that single targets or objects that appear abruptly on a display are rapidly detected in visual search [16]. In addition, if a small number of objects (<5) are precued in some fashion, a user is able to spatially index these objects [17]. Having a spatial index for the selected items allows the items to be tracked simultaneously, causes only the selected items rather than all items to be used in the search space, and allows the rapid enumeration of the selected items. These attentional properties could be used to enhance collaborative interaction not only in a 3D shared space but in a 2D shared space as well. For example, if a collaborator wanted to point something out using a telepointer, the other collaborator's attention would be drawn more quickly if the telepointer were to be set on and off abruptly in a sort of flashing fashion. Another telepointer example could involve the use of precuing. If in a given task it is important that all collaborators be able to keep track of one another's telepointers, having the telepointers precued at the outset would enable the telepointers to be spatially indexed, which would facilitate tracking.

Precuing could be accomplished by having the telepointers flash or marked in an obvious way at the onset.

As I noted above, taking advantage of these attentional properties is not restricted to 3D space. In fact, the studies that were documented [16, 17] were all done using a 2D visual space. It may be the case, therefore, that these properties do not hold in 3D the way that they do in 2D. Further research in this area could be required.

5 Summary

Supporting remote collaboration in 3D presents a fascinating new area of research. The technology required to sustain such collaborative 3D systems will push the technological frontiers. Developing the intricate technology, however, is only half the battle. Significant research will be required in the domain of human-computer interaction. Protocols or metaphors for interacting in 3D spaces will be needed so that this interaction is as comfortable, natural, and fluid as it is in the real 3D world.

This paper represents only the first step in researching 3D interaction protocols and metaphors. It has been an attempt to identify and summarize the relevant literature and to extrapolate from what is known to what is unknown. How to interact in 2D spaces is known. Although it has not been perfected, it has been heavily researched. A second area that has been heavily researched, but is far from perfected, is single-user interaction in 3D. The latter type of interaction is documented primarily in the virtual reality literature. I have attempted to use findings from these two domains of research and speculate on their applicability in 3D multi-user environments.

I have made some speculations, drawn some conclusions and identified some areas that require further research. I will briefly highlight each of these. I speculate that audio, colour and the use of different views will be equally beneficial in 3D as in 2D. Trying to speculate about the use of video is more difficult. When 2D video is used to give a sense of presence while attending to a task supported by a different medium, it is not found to add significantly to the outcome of the collaborative interaction. As such, it seems fairly safe to conclude that using 3D video in a similar scenario would provide little benefit. If video is the sole medium used to support the task domain then it is highly probable that moving the video images from 2D to 3D would enhance the interaction. The one area that would suffer under such an extension, however, is the privacy of the individuals collaborating. The use of 3D video and the balance between privacy and utility is an area that requires further research.

Another area of research that should be consulted in order to develop robust 3D interactions is that of attentional psychology. Research has been done to determine what grabs attention, what holds attention, what facilitates simultaneous attention to multi foci, what the limits of human attention are, and many other attention-related issues. This research, although it may seem peripheral at first glance, should not be ignored. Collaborative interaction in 2D spaces has been under study for at least the last six years and has yet to be perfected. Given the intricacies of interacting in a 3D space, one could speculate that it will be significantly more difficult to perfect than its 2D counterpart. Incorporating research about human interaction with various visual stimuli can only improve and speed along the process of developing 3D interaction protocols. One question that will require further research is: *Does attention differ with 3D stimuli and if so, how?*

The various areas of research literature covered here should not be seen as exhaustive of potentially useful literature for the study of 3D interaction. They should be seen, rather, as some of the key areas that will enable progress to be made in 3D interaction research and that will help uncover issues and other areas of research that require further investigation.

6 References

[1] Jim Hollan and Scott Stornetta. 1992. Beyond being there. *CHI '92 Conference Proceedings*, 119-125.

[2] William Buxton. 1992. Telepresence: Integrating shared task and person spaces. *Graphics Interface Proceedings '92*, 123-129.

[3] Elin Pedersen, Kim McCall, Thomas Moran, and Frank Halasz. 1993. Tivoli: An electronic whiteboard for informal workgroup meetings. *INTERCHI '93 Conference Proceedings*, 391-398.

[4] Saul Greenberg and Ralph Bohnet. 1991. GroupSketch: A multi-user sketchpad for geographically-distributed small groups. *Graphics Interface Proceedings '91*, 207-215.

[5] Ronald Baecker, Dimitrios Nastos, Ilona Posner, and Kelly Mawby. 1993. The user-centred iterative design of collaborative writing software. *INTERCHI '93 Conference Proceedings*, 399-405.

[6] Jin li and Marilyn Mantei. 1992. Working together, virtually. *Graphics Interface Proceedings '92*, 115-122.

[7] William Gaver, Thomas Moran, Allan MacLean, Lennart Lovstrand, Paul Dourish, Kathleen Carter, and William Buxton. 1992. Realizing a video environment: Europarc's Rave system. *CHI '92 Conference Proceedings*, 27-35.

[8] Dan Venolia. 1993. Facile 3D direct manipulation. *INTERCHI '93 Conference Proceedings*, 31-36.

[9] Lennart Fahlen, Olov Stahl, Charles Grant Brown, and Christer Carlsson. 1993. A space based model for user interaction in shared synthetic environment. *INTERCHI '93 Conference Proceedings*, 43-48.

[10] Christer Carlsson and Lennart Fahlen. 1993. Integrated CSCW tools within a shared 3D virtual environment. *INTERCHI '93 Conference Proceedings*, 513.

[11] S.R. Ellis. 1991. Nature and origins of virtual environments: A bibliographic essay. *Computing Systems in Engineering 2*, 4, 341-347.

[12] Michael Deering. 1992. High resolution virtual reality. *SIGGRAPH '92*, (Chicago, July 26-31), 195-202.

[13] Colin Ware, Kevin Arthur, and Kellogg Booth. 1993. Fish tank virtual reality. *INTERCHI '93 Conference Proceedings*, 37-42.

[14] Kevin Arthur, Kellogg Booth, and Colin Ware. 1993. Evaluating 3D task performance for fish tank virtual worlds. 1993. *ACM Transactions on Information Systems 11*, 3, 239-265.

[15] Chris Shaw, Mark Green, Jiandong Lian, and Yunqi Sun. 1993. Decoupled simulation in virtual reality with the MR Toolkit. *ACM Transactions on Information Systems 11*, 3, 287-317.

[16] Steven Yantis and John Jonides. 1984. Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 601-620.

[17] Zenon Pylyshyn. 1994. Some primitive mechanisms of spatial attention. *Cognition 50*, 363-384.

[18] William Gaver, Abigail Sellen, Christian Heath, and Paul Luff. 1993. One is not enough: Multiple views in a media space. *INTERCHI '93*, 335-341. (not summarized)

Acknowledgement

This research has been funded by the Natural Sciences and Engineering Research Council of Canada and the Media and Graphics Interdisciplinary Centre at the University of British Columbia.