

**When Rich Media are Opaque:
Spatial Reference in a 3-D Virtual World**

*Susan C. Herring**

Katy Börner

Maggie B. Swan

[herring; katy; mbswan]@indiana.edu

School of Library and Information Science

Indiana University, Bloomington

* Contact for correspondence

Abstract

Graphical multi-user environments such as 3-D virtual worlds aim to recreate the communicative affordances of face-to-face discourse, based on the assumption that rich media are superior to text-only media for the purposes of most human interaction. However, little research has as yet sought to determine the extent to which face-to-face discourse strategies are used in 3-D virtual worlds, and with what effects. In this paper, spatial reference in one graphical 3-D environment is investigated in the context of a cooperative task that requires subjects in teams to communicate with one another as they navigate through the virtual world. The inexperienced subjects start out employing face-to-face reference strategies, but upon discovering that these do not work the same as in the non-mediated world, adapt in the direction of increased focus and specificity of spatial reference, relying increasingly on linguistic cues and decreasingly on the graphical interface. In light of these findings, we suggest that the rich graphics and multidimensionality of the environment can be misleading: they seduce users into believing that spatiality in the virtual world obeys the same laws as in the physical world, thereby increasing the likelihood of miscommunication, rather than facilitating natural interaction.

Keywords: 3-D, avatars, CMC, discourse, deixis, graphical worlds, media richness, navigation, objects, reference, space, user adaptation, teams, virtual reality.

1. Introduction

As interest in the effects of multi-user computer networks on human communication grows, attention is turning increasingly from text-only modes such as email and chat to media that employ graphical interfaces. In some cases, these media are navigable; that is, in addition to interacting via text messages and graphics, users can move (or direct graphical avatars representing themselves) through the environment, which can be extensive and extensible. Graphical, navigable computer-mediated communication (CMC) environments, or ‘virtual worlds,’ can be two-dimensional or three-dimensional, depending upon whether the navigation space is ‘flat’ (having only two dimensions, height and width), or whether it creates the experiential illusion of depth, the third dimension.

3-D virtual environments are often proposed as training and collaboration spaces that enable users to learn, work and socialize at a distance in ‘rich’ mediated environments that approximate the fullness and interactivity of face-to-face communication (cf. Daft & Lengel, 1984, 1986). For example, 3-D virtual worlds enable movement not just into and out of defined spaces in the interface (which is also possible in 2-D worlds), but around and above virtual objects, including embodiments (‘avatars’) of other participants, allowing for complex shifts in participant perspective, socially-significant spatial positioning, and non-congruent ‘hearing’ and ‘seeing’ ranges. It follows logically that such environments, to the extent that they succeed in recreating the richness of face-to-face environments, should allow users to transfer their face-to-face discourse strategies when communicating online. Specifically, strategies for selecting and addressing interlocutors, choice of discourse topic, and the use and resolution of spatial referring expressions in 3-D graphical virtual worlds might be expected to be similar to those employed when conversing face-to-face. To our knowledge, however, no research has yet attempted to validate or disprove this commonsense expectation.

In the present study, we investigate how novice users respond linguistically to the spatial and navigational properties of a 3-D virtual world known as iUni, which was designed using

Active Worlds technology.^{1[1]} Specifically, we analyze forms of textually-produced spatial reference by participants in a synchronous collaborative activity (an ‘information treasure hunt’) that took place on two occasions in iUni, over a roughly one-hour period each time. Our results show that the 3-D virtual world initially led users to employ face-to-face reference strategies, including the use of deictic terms that require a shared visual context. When the users discovered, however, that spatiality and vision did not function the same as in non-mediated physical reality, they struggled and ultimately adapted in the direction of increased focus and specificity of spatial reference, making their messages more context-independent. In light of these findings, we suggest that the rich graphics and multidimensionality of the environment, given the current level of sophistication of 3-D virtual reality design, may in fact be misleading: they may seduce users into believing that the virtual space obeys the same laws as spatiality in the real world, thereby increasing the likelihood of miscommunication (and potentially, of frustration), rather than facilitating natural interaction, as media richness theories would predict.

In the next section, we survey previous research on media richness and communication in 3-D virtual worlds, with a focus on spatial reference, and describe the virtual environment, iUni, in which the data were collected. The subsequent section lays out the data collection and analysis procedures. The findings are then presented in both qualitative and quantitative terms, to demonstrate how spatial reference in iUni changed over time under conditions of disorientation and disorder. We conclude by considering the theoretical and practical implications of these findings, identifying limitations of the analytical approach, and suggesting directions for future research.

2. Background

2.1. Media richness

The theoretical framework that informs this study is media richness theory, originally articulated as 'information richness theory' by Daft and Lengel (1984, 1986). This theory posits that ‘rich’

communication media, i.e., those that communicate information via multiple channels and provide immediate feedback, enable more complex, socially nuanced communication than ‘lean’ media, which make use of fewer channels and in which feedback is limited or delayed. The prototypical example of a rich medium is face-to-face speech, which communicates through sound, sight, and sometimes touch and smell (if the interlocutors are physically proximate). In contrast, Daft and Lengel give as an example of a lean medium the computer (by which they originally meant number crunching and databases; the attribution was later extended to text-only computer-mediated communication), in as much as it makes use of alphanumeric characters (a narrow subset of visual information), and registers low on the feedback scale in the case of asynchronous modes such as document processing. Telephone conversations (and synchronous modes of text-based CMC such as chat, MUDs and MOOs, etc.) fall along a continuum between these two extremes. Graphical virtual reality environments, although not envisioned by Daft and Lengel in the mid-1980s, would be considered rich according to this framework, as would video conferencing. The relative position of various spoken, computer-mediated, and written media arranged along a continuum from most to least rich is shown in Figure 1.

RICH

Face-to-face communication
Video conferencing/3-D virtual worlds
Audio conferencing/telephone
Synchronous text-based CMC
Asynchronous text-based CMC
Written letters
Formal written documents

LEAN

Figure 1. A continuum of media richness (adapted from Daft and Lengel 1984)

Rich media have advantages relative to lean media. According to Daft and Lengel, lean media are useful for routine, predictable tasks, but richer media such as the telephone or face-to-face communication are superior for complex tasks and tasks requiring ambiguity resolution. Murray (1988)'s study of the medium choices of workers in a large U.S. computer corporation supported this claim: when given the choice of face-to-face, telephone, synchronous chat, and email, workers preferred the richer (spoken) media for more complex (including more interpersonally sensitive) interactions. Rich media also increase social presence and facilitate interpersonal interaction and liking, according to Short, Williams and Christie (1976). Conversely, lean media were found to create interpersonal distance and conflict in early experimental studies by Kiesler, Siegel and McGuire (1984) comparing face-to-face with email and synchronous computer-mediated communication. Finally, people tend to like rich media better than lean media (although in most studies, they like lean media, too), and report high levels of enjoyment and satisfaction even when they do not perform as well on tasks in the rich media (Walther 1999). Perhaps for this reason, many designers focus on creating multimodal systems, aiming to approximate as closely as possible the richness of face-to-face communication, e.g., in video conferencing systems and 3-D virtual worlds. A goal of such systems is to provide users with an authentic context for interaction that feels like a place, rather than a system (Croon Fors and Jakobsson 2000; Harrison and Dourish 1996).

Rich media are popular because, in the best of cases, they can foster a subjective experience of presence or immersion in the mediated environment, in which the medium itself becomes transparent. According to Lombard and Ditton (1997), mediated presence is itself a pleasurable experience, one that can give rise to losing oneself in the "flow" (cf. Csikszentmihalyi 1990), an effect that video games and 360-degree movie screens with wraparound sound aim to create. The perceptual illusion of being co-present with another person can also enhance

sociability (Short et al. 1976). Lombard and Ditton identify a number of features of media that create the illusion of presence, including realism, or the degree to which the mediated environment resembles the real (physical, unmediated) world.

Most virtual worlds incorporate a degree of realism in their graphical representations, such that avatars tend to take human, gendered forms (Kolko 1999; McDonough 1999), and environments are comprised of buildings, paths, and realistic or semi-realistic landscapes.^{2[2]} A common claim for the intuitiveness and superiority of 3-D virtual worlds is that they exploit and build upon users' everyday skills to understand spatial arrangements and to interact within the physical world (Benford, Bowers, Fahlen, Mariani and Rodden 1994). In the present study, we are interested in the ways in which the iUni environment is realistic (or not), and how its resemblance to the unmediated physical world influences users' discourse choices.

2.2. Communication in rich media

Virtual worlds, like other forms of collaborative multimedia, are intended to facilitate human communication and interaction. As yet, however, there have been few linguistic studies of communication in rich media. Most language-focused CMC research has concentrated on text-only modes such as email (Baron 1998; Cho In press), discussion groups (Baym 1996; Herring 1996), chat (Paolillo 2001; Werry 1996), social MUDs (Cherny 1999), and more recently, instant messaging and text messaging via mobile phones (Hård af Segerstad 2002).^{3[3]} What research exists on the effects of interactive multimedia environments on discourse has been carried out mostly through experimental studies in the field of Computer-Supported Cooperative Work (CSCW).

Most relevant to the present study is research on video-mediated communication. Video-mediated communication, also known as desktop videoconferencing, involves two or more people in different locations conversing via video cameras and microphones which record moving, photo-realistic images (typically of 'talking heads') and speech which are transmitted

synchronously over networked computers.^{4[4]} When the quality of the transmission is high, interlocutors can see one another's facial expressions and gestures, as well as hear prosodic information in speech. Video-mediated communication is thus potentially very rich; Goddard (1995; cited in Fischer and Tenbrink 2002) claims that video conferencing is "the next-best medium for interaction where face to face contact is not feasible." CSCW research on desktop videoconferencing has investigated turn-taking (Sellen 1992, 1995) and effect of image size and perceptual closeness on overall levels of interactivity (Fischer and Tenbrink 2002; Grayson and Coventry 1998). In one of the few studies conducted by linguists, Yates and Graddol (1996) found that the visual information available in video-mediated communication influenced topic of conversation in a comparison across media: interlocutors communicating via CUseeMe talked more about physical appearances than interlocutors in either text-only CMC or face-to-face communication. Video conferencing has also been found to cause confusion in talk about objects in shared virtual space, leading interlocutors to modify their strategies of spatial reference (Barnard, May and Salber 1996). Despite distortions introduced by the video conferencing system, "users treat video interactions like face-to-face conversations" (Heath and Luff 1991; cited in O'Malley, Langton, Anderson, Doherty-Sneddon and Bruce 1996), even when it leads them to make task-related errors. Similar findings have been observed for virtual worlds.

2.2.1. Communication in virtual worlds

Graphical virtual worlds facilitate multi-participant interaction through the use of graphics and, in some cases, animation, video, and sound. Users not only exchange words and see images of one another, they are able to navigate through the environment and, in the case of 3-D virtual worlds, interact with objects (including other avatars, which are technically objects) in the virtual environment. Even more than video conferencing, such systems are claimed to foster telepresence, or immersion in the mediated environment. Desktop systems enable this experience to take place collaboratively via the Internet, without the encumbrance (and disorienting effects)

of special headsets, glasses, and/or gloves typically required by immersive 3-D systems (Robertson, Card and Mackinlay 1993).^{5[5]}

At the same time, most graphical worlds available today are not as rich as they could be, nor as rich as video conferencing in certain respects. As Steuer (1995) notes, "media systems that allow individuals to interact with each other in natural ways within virtual environments are not yet common, nor are systems that can represent the seemingly infinite range of sensory raw materials present in the real world." Although systems are currently being developed that allow users to communicate through speech, most require that communication take place through the exchange of synchronous text messages; thus prosodic information is lacking. Moreover, graphical avatars are severely limited in the range of meanings they can express through movement, gesture, and facial expression, in comparison with the human body (and especially the human face). Manninen and Pirkola (1997) compared the features of 15 virtual worlds, classified into three types: text-only (MUDs), 2-D, and 3-D. The sample included worlds designed for education, work, social chat, and recreation (games). They found that despite the richer interface of the 3-D worlds, the three world types were equally 'realistic', and the text worlds actually allowed users the greatest expressive potential.^{6[6]} They concluded that "[c]urrently, there seems to be no perfect, or even adequate, multi-user virtual world" (Manninen and Pirkola 1997).

Claims regarding the effects of graphical virtual worlds on communication are also mixed. First-person ethnographic accounts of user interaction in 2-D and 3-D graphical environments such as The Palace (2-D) and Active Worlds (3-D) have reported high levels of sociability and 'community-like' behaviors (Eep2 2000; McClellan 1996; Scannell 1999; Suler 1996). Similarly, Neal (1997) reports that having students in a distance education course meet in the 2 1/2-D^{7[7]} environment WorldsAway resulted in informal and humorous communication, and "enhanced community-building in the class." Contrary to the idea that virtual reality creates ideal social worlds, however, McClellan (1996) reports on the existence of antisocial gang behaviors, including vandalism, in Alpha World (the original Active World), and Allwood and Schroeder

(2000) find that speakers of languages other than English are marginalized in mixed-ethnicity Active Worlds.

A smaller number of studies have investigated how users engage with the semiotic resources of the graphical environment itself. Krikorian, Lee, Chock and Harms (1999) found that avatar distance is socially meaningful in The Palace: in an experimental study, distance significantly influenced subjects' social liking, with close proximity of two avatars in the interface giving rise to intimacy and/or a sense of “crowding”, depending on the personality characteristics of the subject. Naper's (2001, In press) analysis of natural group interaction in Patagonia, a Scandinavian Active World, reveals that avatars themselves can be an interactional resource. For example, although avatars in Patagonia have fixed facial expressions, a user can change avatars to express different moods (see also Suler 1996, for The Palace). Another example of avatar-based communication is the exchange of avatar heads in WorldsAway to signal friendship or romantic involvement (Scannell 1999; see also Neal 1997).

Manovich (1996) writes that “[v]irtual worlds keep reminding us about their artificiality, incompleteness, and constructedness.” A removable head is obviously not a feature of a real human being. Even in 3-D virtual worlds designed for serious purposes such as work or training, the illusion of reality is easily disturbed, for example by the lag that typically results when a scene is automatically redrawn after an avatar moves or shifts perspective, giving rise to ‘jagged’ visual transitions. However, as Mitchell (1995) notes in describing a virtual reality environment (SIMNET) designed to train soldiers:

An interesting lesson learned from SIMNET was that the users felt immersed even though the display had a rather poor resolution. Soldiers still found the system thrilling to use, and they “saw through the technology”, as one expert described. The 3D world was comprehensible and consistent, and the user’s imagination could fill in the missing resolution.

The same principle—imagination filling in the gaps—accounts for user experiences of immersion in text-based virtual worlds (Manninen and Pirkola 1997). On the one hand, it shows that users are forgiving of the limitations of the technology. On the other hand, in cooperating with a medium's aspirations, rather than its actual affordances, users may inadvertently adopt face-to-face strategies that are inappropriate for the medium. This effect can be seen in research on mediated spatial reference.

2.3. Spatial reference

Interactive multimedia are designed to support cooperation among remote users. They do this, in part, by allowing users to refer to objects and features visible in a shared graphical interface. Reference to common objects is especially important in video conferencing systems and virtual worlds used to support group work processes, such as the design of automobiles.^{8[8]} In 3-D virtual worlds, in addition, collaborators must sometimes refer to objects that are not simultaneously visible to all participants, and direct others to them. The challenge common to users in these situations is to communicate unambiguously about the location of objects in virtual space to interlocutors who are not physically co-present.

If collaborative multimedia interfaces were transparent, it should be a simple matter to transfer face-to-face strategies of spatial reference to multimedia environments. For example, deixis should work basically the same way in the mediated world as in the unmediated world. Deixis is reference by means of words (such as 'this/that', 'here', 'next to') and/or gestures (such as pointing or gaze direction) that are dependent on context for their interpretation (Fillmore 1997; Levinson 1983; Lyons 1977). In order for spatial deixis to communicate successfully, interlocutors need to have visual access to a common *context*, and share (or be able to interpret) one another's visual *perspective*. More fundamentally yet, interlocutors need to be aware of their physical *orientation* with respect to one another (whether they are face-to-face, next to one another, back-to-front, back-to-back, etc.).

In fact, problems sometimes arise in locating objects in space in collaborative multimedia that point out ways in which the interfaces are not at all transparent. In the case of video conferencing, confusion arises when the video camera is located in front of the user (e.g., on top of the user's computer monitor) to give a 'face-to-face' perspective. The image that is transmitted to the remote interlocutor is reversed relative to the physical reality; thus when one person points to an object within the interface to his left, the other person sees the gesture as pointing to a (different) object on her right. Barnard et al. (1996) conducted an experiment in which novice subjects viewed a video instructing them to identify objects visible through the interface. Subjects made far fewer errors when the instructor referred to the objects with explicit linguistic forms, than when referring forms were accompanied by deictic gestures (which were, of course, reversed). It is noteworthy that in the latter condition, the subjects relied on the gestural indicators over the linguistic cues, just as they might have done in a face-to-face situation. The authors conclude:

Where the participants are implicitly encouraged to believe that they are in a shared workspace, they may all too easily adopt a communicative register allowing them to believe in physical co-presence and therefore assume that a gesture will resolve the ambiguity as it would in the physically co-present setting. (Barnard et al. 1996)

This recalls the observation of O'Malley et al. (1996) that users treat video conferencing like spoken interaction with regard to turn-taking, even when the system interferes with normal turn-taking.

In graphical virtual worlds, spatial reference is complicated by the fact that users who move about must constantly adjust to shifts in perspective. As Darken (1995) points out, every time one moves in a virtual world, one must re-orient oneself to the environment and others in it, a task rendered more difficult by the fact that one's range of vision, especially in first person perspective, is limited. Darken makes the following observations about the difficulties of

'wayfinding' in virtual environments:

Virtual world navigators may wander aimlessly when attempting to find a place for the first time. They may then have difficulty relocating places recently visited. They are often unable to grasp the overall topological structure of the space. Any time an environment encompasses more space than can possibly be viewed from a single vantage point, these problems will occur. (Darken 1995)

Darken (1995) and Darken and Sibert (1996) found that novice users in an experimental study of wayfinding in a 3-D virtual world tended to employ navigation strategies, such as 'dead reckoning',^{9[9]} carried over from the physical world, even though the strategies did not work the same way or produce the desired outcomes.

Few studies have as yet focused on talk about objects and their location in 3-D virtual worlds. In an experiment that required pairs of subjects to communicate about arranging furniture in a desktop collaborative virtual environment, Hindmarsh, Fraser, Heath, Benford and Greenhalgh (2000) found "problems due to 'fragmented' views of embodiments in relation to shared objects[, ...] and difficulties in understanding others' perspectives". Communication in the environment could occur via speech (audio) and avatar pointing. The problems arose from the fact that in order to interpret a pointing gesture by another avatar, a subject had to be able to see both the avatar and the piece of furniture referred to. However, as avatars only had a 55 degree field of vision (as compared to about 180 degrees that a normal human being can see in an open space),^{10[10]} the pointing avatar or the piece of furniture, or both, often fell outside the subject's field of vision. The achievement of shared reference was further confounded by the difficulty of knowing what was in another avatar's field of vision when it referred to 'this' or 'that'. To compensate for these problems and complete the task, subjects gave explicit spoken accounts of their avatars' actions, rather than relying on deixis alone.

This research calls into question the assumption that rich media enable the direct transfer

of face-to-face strategies as regards spatial reference, suggesting rather that users must accommodate linguistically to the properties of the computer-mediated systems, or risk miscommunication (and reduced task performance). The results of the present investigation provide further evidence in support of this view. In addition, they show how users adapt linguistically over a short period of time, shedding light on the processes of, and underlying motivations for, communicative accommodation to graphical virtual worlds.

3. Research Site

The site analyzed in this study, iUni (information Universe; Börner 2001), is a world in EduVerse, the educational analog of the recreationally-focused Active Worlds universe. The Active Worlds (AW) system was developed by Ron Britvich for Knowledge Adventure Worlds, and made publicly available on the Internet in 1995. Active Worlds provides a synchronous, text-based chat facility embedded within a three-dimensional graphical environment, which users—represented by humanoid graphical avatars^{11[11]}—can navigate by walking, flying, or teleporting from location to location.

The iUni world was designed in fall 2000 by graduate students taking the second author's User Interface Design course at the Indiana University (IU) School of Library and Information Science. Students collaborated with IU faculty on the design of customized 'Learning Environments' for different classes. These include an educational 'Natural Disaster Area', a 'Science House', a 'Quest Atlantis' portal to different theme parks for children, an 'Art Café', and a 'Virtual Collaboration Area' intended to provide a round-the-clock, worldwide accessible meeting space for business professionals.^{12[12]}

As Figure 2 shows, the iUni interface contains three main windows: a 3-D graphics window populated by avatars (the 'world' itself), a chat window, and a Web browser window. At the top are a menu bar and a toolbar for controlling avatar actions. Users can navigate in three dimensions, move their mouse pointer over an object to bring up its description, click on 3-D

objects to display the corresponding Web page in the right Web frame of the EduVerse browser, or teleport to other places within the world. The Web browser maintains a history of visited places and Web pages so that the user can return to previous locations.

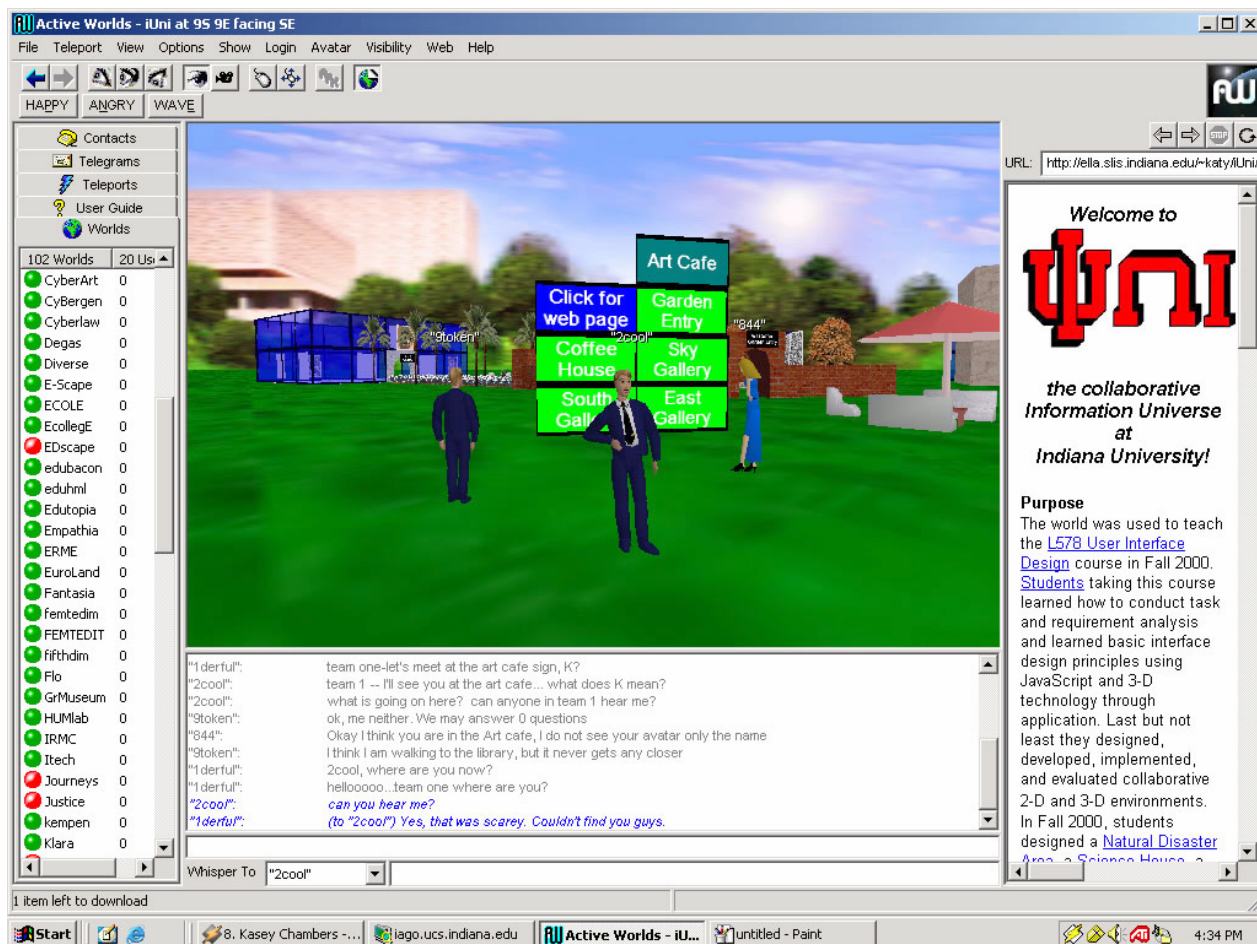


Figure 2. The iUni interface

Visitors to the world have a choice of two avatars, one male and one female (the default avatar is male); both types are represented in Figure 2.^{13[13]} The avatars available at the time of our study were capable of a limited set of preprogrammed motions ('dance', 'wave', 'angry', 'happy') that could be activated through the menu bar, in addition to being able to navigate by 'walking', 'flying', or 'teleporting' from one location to another. Obviously, flying and teleporting

are not navigation possibilities normally available to individuals in the real world.

Also unlike the real world, users have a choice of first person or third person perspective. In the first person perspective, the user 'sees' the world directly, as though through the eyes of her avatar. In the third person perspective, the user sees her avatar against the backdrop of the environment as though it were an actor (or a puppet) on a stage. The 'third person perspective' provides a slightly elevated view down upon the avatar and its surroundings from behind. In addition, the 'flying' navigational technique provides an aerial view and allows the avatar to elevate or float above its surroundings. One can fly at a continuous range of heights in first or third person mode. In a study by McMahon and Börner (Under consideration), subjects indicated a preference for flying and third person views in a task that required them to locate a web page thumbnail. In Active Worlds, due to an inconsistency in the design of the environment, users have a narrower range of vision in the third person perspective, and objects appear closer, than in first person perspective, as can be seen by comparing Figures 3a and 3b.



Figure 3a. First person perspective



Figure 3b. Third person perspective

In addition to navigation, iUni is designed to facilitate communication and interaction. Users communicate with one another by typing text messages into a message buffer at the bottom of the chat window. Their messages appear both above the heads of their avatars, and in sequential order following their user name in the chat window, as in other forms of synchronous chat. Moreover, visitors to iUni can 'whisper' to other visitors (send a private message) regardless of where their avatars are located in the world. 'Whispered' messages can be identified in the chat window by the presence of the name of the addressee in parentheses after the name of the sender, and by the special color (blue) of the text. (See the last two lines in the chat window in Figure 2.)

Only those avatars located within a certain distance of one another can read one another's

messages in the chat window; if they move too far away, they go out of ‘hearing’ (reading) range. In Active Worlds, the hearing range is a distance that would correspond to 120 meters in the physical world.^{14[14]} While this distance is greater than normal human hearing range for unamplified speech in the physical world, it is less than the ‘hearing’ range in most chat rooms, which make all messages posted within a chat space equally ‘audible’. Conversely, chat log files in iUni only show and collect utterances of users’ avatars that are within a certain range. Confusion sometimes arises in iUni when avatars move out of chatting range. Even at close range, moreover, there is no ‘spatialized sound’ available to indicate the spatial source of typed chat text, making it a challenge at times to locate discussion partners.

Seeing one's discussion partners can also be tricky. As in other graphical 3-D virtual worlds, an avatar's visual range is restricted to less than 60 degrees. Moreover, unlike real human beings, iUni avatars are not able to move their eyes, turn their heads, or swivel at the waist to scan the environment visually. Rather, users must turn their entire avatars to find out who is talking to them. The need to turn one's avatar whenever one wishes to look around, combined with the narrow visual range, means that one can easily ‘lose’ another avatar from sight, even when it is located close by. To ease identification, the Active Worlds browser displays the user name above each avatar in a near-constant size. When avatars are far away, their names tend to appear jumbled together on top of one another (this can be seen in Figure 3a), but one can still identify groups of avatars from afar. However, a single avatar can easily be overlooked and missed if it is distant or hidden behind a textured wall or other object.

In its navigational and communicative affordances, therefore, iUni is not entirely like the real world. As in physical space, movement in three dimensions is possible and distance matters for purposes of communication. Unlike in the real world, however, avatars have superhuman navigational abilities and 'whispering' transcends distance. Moreover, iUni avatars are more limited than humans in how they perceive the environment, and the conditions under which ‘seeing’ and ‘hearing’ ranges differ are not the same as in face-to-face communication.

iUni also differs from textual chat environments such as MUDs and chatrooms, although it shares features (such as the 'whisper' command) with them as well. Obvious differences are the graphical interface and the embodiment of the user in the form of an avatar, which enrich the system (Naper In press). Moreover, in the interests of realism, iUni does not allow users to broadcast messages to everyone present in the environment, nor facilitate users determining the location of other users (i.e., at the time of our study, it was not possible to obtain a list or map of where users were located), other than through whispering to them and asking "where are you?" This latter feature made co-located conversation somewhat of a challenge, especially in the second data set in our study, in which the 'whisper' command was not available.^{15[15]} As the evidence presented below attests, these features of the 3-D virtual environment influenced users' communicative behavior.

4. Methodology

4.1. Study design

In order to investigate the effects of the iUni environment on spatial reference, we designed a semi-experimental study in which 27 students in two master's level information science classes at Indiana University participated in an information treasure hunt in iUni. The treasure hunts were administered near the end of the 2001 and 2002 spring semesters by the first and second authors, who were also the instructors of the classes. Eight subjects participated in the first treasure hunt, and 19 in the second; participant identities were masked by the use of self-selected nicknames based on randomly-assigned numbers. The treasure hunt required the subjects to cooperate in teams of two to four people to find and share information about the virtual environment. Subjects were asked to locate 10 specific pieces of information (see Appendix).

The subjects were seated together in a computer lab; each had access to an IBM PC-compatible computer. Before starting the experiment, subjects read a one-page description of the study and signed consent forms to allow records of their interaction to be analyzed for research

purposes. A pre-test questionnaire was then administered to gather demographic data about the subjects such as age, gender, native language, handedness, average number of hours spent on the computer and the Internet per week, and the subjects' prior experience with 3-D virtual worlds. The subjects ranged in age from 21 to 46 years old; most of the students in the first class were female, whereas those in the second class were roughly evenly divided between females and males. A number of the students in the second class were non-native speakers of English. Only three students (all in the second class) had ever visited a 3-D multi-user virtual environment before, although all but two (also in the second class) had been in text-only chat rooms.

After completing the pre-test, subjects were instructed to access iUni by starting the EduVerse Browser from the desktop and logging in as guests. Subjects were then provided with a brief introduction to the environment and navigational instructions, and given five minutes to explore the different navigation and interaction possibilities in iUni before beginning the main task, which was to search collectively for the answers to the 10 questions about the world. The instructions recommended that the subjects "meet with [their] team members", "decide on a strategy for how to find the information quickly" and "go hunt for information and exchange results". Each treasure hunt was timed, and lasted for approximately one hour. Although the subjects were in the same room and could have spoken or gestured to one another, they made little attempt to do so, perhaps because the course instructor was also present. Communication about the treasure hunt task took place exclusively through the iUni interface.

Our previous observations of 3-D virtual worlds had indicated that people tend to keep their avatars located in one place while conversing, and that they tend not to converse while navigating the environment. The goal of the information treasure hunt was to ensure that subjects would move about the environment *and* communicate with one another, thereby making use of both the navigational and the communicative affordances of the system. In addition, the treasure hunt task required subjects to locate and refer to objects within the environment, necessitating frequent use of forms of spatial reference. Logs of all textual interactions were collected.^{16[16]} The

first treasure hunt produced 228 text messages, and the second, 658 messages. We analyzed all of the first set of messages, and the first, middle and last 100 messages of the second set, for a total of 528 messages.

4.2. Analytical methods

The text messages were analyzed both qualitatively and quantitatively using methods of computer-mediated discourse analysis (Herring 2003). This involved establishing coding categories for linguistic and behavioral phenomena and coding and counting all instances of those phenomena in the data, taking into account the discourse context. The results were then subjected to statistical analysis using GLMstat, a regression analysis program for the Macintosh computer, and interpreted in relation to the properties of the medium and the social situation.

Five sets of discourse phenomena were identified and counted, as described below:

I. Spatial Reference

Forms of spatial reference were classified into three types, according to the degree to which they depend on the shared visual context for their interpretation.

1) *Deictic* forms of reference are entirely context dependent. They include items such as 'here' and 'this', whose meaning is shifting and non-unique. (That is, the meaning of 'here' shifts according to the location of the speaking ego; it follows that there is not a single place named 'here' in the environment.) Deictic forms were further subdivided on structural grounds into a) adverbs (e.g., 'here', 'back') and prepositions (optionally followed by first or second person pronouns, e.g., 'under', 'near me') and b) verbs (e.g., 'come', 'go', 'look at', 'wait up') (cf. Fillmore 1982).

2) *Fixed, non-unique* forms of reference are partially context dependent. They include such

expressions as 'near the trees' and 'in the air'. In this category, locational prepositional phrases, which may include deictic prepositions such as 'near', are anchored by the inclusion of a non-deictic noun phrase such as 'the trees'. However, in as much as there are multiple clusters of trees in iUni, 'near the trees' is not a unique place.

3) *Fixed, unique* forms of reference are context independent. An example is the expression 'near the art café sign'. Although this expression includes a deictic word ('near'), it is anchored in relation to a non-deictic noun phrase ('the art café sign'), and hence its reference is fixed. Since there is only one art café sign in iUni, the reference is also unique.

There are more constraints on the successful use of strategy (1) than on the higher numbered strategies. Strategy 3) will communicate successfully if both speaker and addressee(s) merely believe that the referred-to entity exists. In strategy (2), in addition, speaker and addressee(s) should be in the same general vicinity, even if they can not see one another (this is so that they will orient to the same cluster of 'trees', for example). For strategy (1), deixis, to succeed, however, speaker and addressee(s) must see one another's avatars, and possibly also note the other avatar's orientation. For example, the utterance "There, you're right next to me", directed by one of the subjects in our study to another avatar, only communicates the speaker's intended meaning if the addressee can see the speaker's avatar. An utterance such as "It's in front of me" only communicates if the addressee can also see the direction in which the body of the speaker's avatar is facing.

II. Address

As a precondition for successful communication, subjects in the virtual task must first locate and direct messages to their intended addressees—in this case, their teammates. We distinguished in our analysis among four address strategies: 1) Unspecified addressee (e.g., 'Where should we

start?'); 2) Team as addressee (e.g., 'hi team 2'); 3) Individual addressee (e.g., 'hi 5 star'); and 4) Private communication to an individual via the 'whisper' command (e.g., '(to "5star" where r u)').^{17[17]} These strategies are numbered from least to most specific. Targeting a user specifically can help to facilitate coherent communication in an otherwise confusing environment.

III. (Dis)orientation

As Darken (1995) and others have observed, it is common for first-time visitors to 3-D virtual worlds to experience spatial disorientation. We coded utterances that explicitly referred to subjects' cognitive state in relation to the mediated environment, classified into three categories: 1) Expression of disorientation (e.g., 'I am lost'); 2) Expression of emergent understanding (e.g., 'I guess I was too far away'); and 3) Expression of mastery (e.g., 'I'm finally getting the hang of this'). Subjects who understand how the virtual world works may be more likely to employ linguistic strategies appropriate to the mediated environment.

The above categories measure users' adaptation to the virtual environment as it is reflected in their discourse. In addition, we devised two sets of codes to control for the effects of the task on subjects' communication, also based on the content of their utterances. These codes, which are described below, are concerned with task focus and adherence to the task schema suggested in the experimental instructions.

IV. Task Focus

Previous research has noted that graphical worlds tend to encourage playful behavior (e.g., Neal 1997). We therefore coded the function of each utterance in the sample into one of three general categories: 1) *Off-task* and unrelated to the virtual world (e.g., 'hello everyone'); 2) *Off-task and oriented to the virtual world* (e.g., 'look at me I'm flying!!'), and 3) *Task-focused* (e.g., 'Should we tackle question 1?'). Especially interesting for the purposes of the present study are off-task utterances that relate to the virtual world. These can be taken as an indicator of the extent to

which the virtual world is distracting to novice users.

V. Task Schema

For those utterances coded as IV.3 ("Task-focused"), we assigned a code to indicate the nature of the task-related activity. Five activities were identified, as suggested by the instructions given at the beginning of the experiment: 1) Talk about team membership (e.g., 'we are on the same team'); 2) Talk about setting meeting places (e.g., 'Team 1, where should we meet?'); 3) Planning/coordinating team activities (e.g., 'I will try to find student residences, you go for number 3'); 4) Talk about going places to seek information (e.g., 'quest atlantis is on the way to the library'); and 5) Collecting/reporting answers (e.g., 'hey, have you found any info?'). These activities constitute a logical sequence, or schema, through which subjects might reasonably progress to complete the task.

The five coding categories and 22 individual codes described above were applied to all 528 utterances in the iUni chat log sample.

The second part of the analysis involved dividing each data set into subsets according to team membership. There were eight teams: three in the first treasure hunt, and five in the second. Correlations were identified between the 22 coding categories and success in the information treasure hunt, as measured by the number of answers reported at the end of the chat logs.^{18[18]} We hypothesized that use of the lower numbered (non-specific; disoriented; off-topic) discourse behaviors would be less successful than use of the higher numbered (specific; oriented; task-focused) discourse behaviors, in two respects. Specifically, we hypothesized that:

- H1. As they adjust to the system, subjects will exhibit more of the higher numbered behaviors over time. (Individual Adaptation)
- H2. Successful teams will exhibit more of the higher numbered behaviors than unsuccessful teams. (Team Differences)

5. Findings

5.1. User adaptation over time to the environment

The graphical 3-D environment proved to be disorienting and distracting. A high proportion of subjects' utterances were about the environment itself, unrelated to the essential requirements of the task, and remarks indicating disorientation (and sometimes dismay or frustration) were not uncommon. Moreover, the information treasure hunt task proved to be so difficult that no team was able to complete it entirely, and several teams found no answers at all. Despite this, as observed in previous studies, the subjects seemed to enjoy the experience of 'trying out' the virtual world, and playful interaction was evident, especially in the second group. Moreover, despite the impression that a casual observer might have had of near-chaos throughout the experiment, the novice users showed significant change over the course of the experiment towards increased referential specificity. The subjects' linguistic modifications can be understood as adaptive to the properties of the environment, in support of the Adaptation hypothesis. Significantly, as in the studies described earlier, the visual constraints of the environment rendered the use of face-to-face reference strategies problematic, requiring a shift to more context-independent strategies.

Evidence in support of these claims is presented below. Since experimental trial did not prove to be a significant factor for any of the five coding categories in the statistical analysis, the results for the two trials are combined unless otherwise indicated.

5.1.1. Spatial reference

Spatial reference was frequent in the subjects' communication, regardless of the activity in which they were engaged. Deictic forms of reference were the most frequent, despite the fact that they often failed to communicate their intended meaning. An example of unsuccessful use of deixis is given in (1) below, from the second experimental trial. Teammates 09 and 11 have just located

one another, and another teammate, 12rp, sees them. However, 09 and 11 apparently do not see 12rp's avatar, rendering uninterpretable his deictic verb 'follow' in line 115. It is not until line 148, when 09 has finally come face-to-face with 12rp, that the command to 'follow' can be meaningfully interpreted. (This sample is taken from subject 09's log. Communication by other teams that was interleaved in the sample but in which no member of 09's team participated has been removed, to simplify the presentation.)

- (1) 102 "09" said 'hi 11'
- 105 "09" said 'u are in my team'
- 109 "11" said 'hey 10 what's ur name'
- 115 "12rp" said 'hey **follow** me 9 and 11'
- 118 "09" said 'but we have to locate 10 and 12'
- 120 "10" said 'hi 11 can we exchange name?'
- 122 "09" said '10 and 12'
- 124 "09" said 'locate them'
- 125 "12rp" said 'I AM **RIGHT HERE!!!!!!!!!!!!!!!!!!!!!!!!!!!! FOLLOW ME**'
- 128 "12rp": hellooooooooooooooooooooooooooooooooooooo
- 129 "09" said 'wiat'
- 131 "09" said 'wait'
- 132 "10" said '12 how to **follow** you? why I can not move?'
- 133 "09" said 'you wait **here**'
- 134 "12rp" said '**look left**'
- 135 "09" said 'and let me search for them'
- 137 "09" said 'will that be okay'
- 140 "12rp" said 'no stop'
- 144 "09" said 'anand [12rp's name]'
- 145 "09" said '12'
- 147 "09" said 'hi 12'
- 148 "12rp" said '**follow me**'

12rp attempts several times to attract his teammates' attention, including by 'shouting' in all caps in line 125 'I AM RIGHT HERE!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!' and repeating the final vowel of a vocative 'hello' in line 128, as if calling from a distance. However, since the others (except possibly 10) do not see his avatar, the deictic expression 'right here' is as uninterpretable as 'follow', in that both require the addressee to know where the speaker's avatar is located. Nor does 'shouting' or drawing out a final syllable help to attract attention to the location of his avatar, since chat messages appear in temporal (rather than spatial) order in the chat window, and if one does not see an avatar, one may not see the text of a message that appears above the avatar when it 'speaks' in the graphical window. 12rp then tries instructing 09 (or possibly 10) to 'look left' (to see his avatar), but gets no response. Meanwhile, 09 tells 11 (?) to 'wait here' and sets off in search of 10 and 12. Only then does he turn and come face-to-face with 12, who (along with 10) has been nearby all along. This section of discourse includes several deictic expressions as the participants attempt to locate one another and describe their own location.

The use of deixis increases throughout the treasure hunt ($p=.043$), contrary to our hypothesis that non-specific reference would decrease as users adapted to the environment. Conversely, in support of the Adaptation hypothesis, subjects tended to increase their use of fixed, non-specific reference over time, and significantly increased their use of fixed, specific spatial reference ($p=.008$). Example (2), from the first experimental trial, illustrates the use of fixed, non-specific reference; example (3) illustrates the use of fixed, specific reference. (Interleaved communication among other teams not relevant to each exchange has been omitted.)

- (2) 70 "fortune": team 2 do you have maps?
- 74 "sixtet": (to "fortune") yes, I am **by the pictures**?
- 75 "sixtet": **by the pictures**
- 77 "fortune": what are the pics of?
- 78 "5star": no idea

In the above exchange, sixtet tries to indicate her location to her teammates fortune and 5star. Her response is more specific than if she had made use of a deictic expression such as 'I am over here'. However, the repeated expression 'by the pictures', while anchored to an immovable feature of the virtual environment, does not uniquely identify her position, since there are several displays of 'pictures' in the team's immediate vicinity. Fortune attempts to elicit additional identifying information by asking 'what are the pics of?' 5star's response, 'no idea', further indicates that sixtet's description was insufficient.

In example (3), in contrast, 1derful and 2cool communicate effectively about an intended meeting place:

- (3) 90 "1derful": team one-let's meet **by the art café sign**, K?
91 "2cool": team one -- I'll see you **at the art café** ... what does K mean?

Although the expression 'by the art café sign' has the same formal structure as 'by the pictures'—both are definite, referential noun phrases preceded by the preposition 'by'—, 'the art café sign' refers to a unique object within the virtual world, and thus its reference is unambiguous. Indeed, the only confusion that arises in this exchange is caused by 1derful's use of the abbreviation 'K' (to mean 'okay'), presumably because of its uncharacteristic capitalization.^{19[19]}

Spatial deixis is a typical face-to-face reference strategy. Context-independent reference, conversely, is characteristic of situations in which the interlocutors do not share physical space, and where explicitness is required, such as in formal writing. That deictic forms of reference persisted and in fact increased over the course of the experiment, even though they regularly gave rise to misunderstanding, can be taken as evidence that the environment encourages users to interact as though they were face-to-face. That some subjects adopted more 'written-like,' context-independent reference strategies over time suggests that they realized that they needed to be more

explicit in order to communicate about locations in the virtual world.

5.1.2. Address

There is a clear tendency in the first experimental trial for subjects to shift over time from the use of no address forms to increasingly targeted forms of address, settling on the 'whisper' command as the preferred option. This can also be seen as a shift from a face-to-face strategy—we do not normally mention our interlocutors' names in each utterance when they are in front of us—to a written strategy—in this case, one characteristic of text-only chat rooms (Werry 1996).^{20[20]} Specific forms of address also facilitate the treasure hunt task, in that they help subjects first to locate their team members, then to share information about the treasure hunt privately with them. The chat logs suggest that the subjects in the second experimental trial, who did not have use of the 'whisper' command, had less focused interaction, and experienced greater difficulty in completing the task, as a result. In example (4), subjects from the second trial are attempting to share the information they have gathered at the end of the treasure hunt. 09, 10, and 12rp are teammates; 02 is in a different team; and 5a is in yet another team. 13-16 comprise a team, as do 17mom, 18, and 19. All are within 'hearing' (although not necessarily seeing) range of one another.

- (4) 253 "09" said 'hi'
254 "12rp" said 'hi did you get any answers'
255 "09" said 'so to how many questions u get the anser'
256 "02": 3a. how are you flying?
257 "09" said 'iyup'
258 "5a" said 'Hi, 907'
259 "12rp" said '8 9 and 10'
260 "12rp" said 'there probably not right though'
261 "09" said 'I got for 6th 7th and 7th'
262 "12rp" said '#10 don't know what's going on'

263 "5a" said 'I found 5 areas in art café'
264 "09" said 'I got answers for 6th 7th and 8'
265 "02" said 'you all wanna share answers? =)'
266 "09" said 'what about u'
267 "12rp" said '8, 9 , and 10 for the 100th time'
268 "09" said 'to how many questions you have the answers'
269 "10" said 'how to go inside room?'
270 "10" said 'I didn't get any'
271 "09" said 'please give me ur answers'
272 "18": mom
273 "12rp" said '#10 we are over here'
274 "09" said 'give me ur ansers'
275 "18": give me answer
276 "09" said 'give me ur answers'
277 "2w" said 'STOP EXPERIMENT!!!!!!'
278 "18": ha
279 "18": ?
280 "13": 13-16..... any success?????
281 "012rp" said 'then everyone will know our answers'
282 "17mom": #8 is 4, #9 is 4
283 "14": none here
284 "15": none here too
285 "09" said 'give me ur answers'
286 "09" said 'pleseeeeeeeeeeeeeeeeeeeee'
287 "13": i found the art cafe.... at least answered one ques!
288 "09" said 'pleaseeeeeeeeeeeeeeeeeeeee'

This ending is considerably more disorderly than the ending of the first experiment, in which subjects made almost exclusive use of the whisper command to report their findings privately to their teammates. Most messages in this sample are *not* addressed to a specific addressee, despite

containing information that is not intended for other teams. The coordination attempts of the team comprised of 09, 10, and 12rp are particularly unsuccessful. 12rp interprets 09's messages to be addressed to him, but 09 apparently doesn't think 12rp is addressing him, or doesn't 'hear' 12rp, because 09 repeats his question 'to how many questions you have the answers' several times, even after 12rp answers it. It is also not clear whom 10 is addressing when she asks 'how to go inside room?' in line 269; her question goes unanswered. That subjects in the second trial feel constrained by the lack of the whisper command is poignantly revealed in line 281, when 12rp hesitates to share his answers because 'then everyone will know our answers'. He says nothing despite 09's repeated pleas, and the experiment ends without the team having completed the final phase of the task.

Despite the difficulties experienced in the second trial, all of the results for specificity of address are in the predicted direction. Over time, subjects in both trials use fewer messages which lack an explicit addressee ($p=.0052$). They also tend to use fewer messages addressed to their team as a whole, and more messages addressed to individuals by name. Use of the 'whisper' command in the first trial also tends to increase over time. These results support the Adaptation hypothesis.

5.1.3. (Dis)orientation

Subjects express considerable disorientation throughout the task. Some examples of utterances counted as disorientation are given in (5).

- (5) 16 "18" said 'where is my avita [avatar]?'
50 "16" said 'which animal [i.e., avatar] is mine?'
104 "15" said 'where am I?'
246 "13" said 'we are lost!'
184 "1derful": Where the heck are we?
48 "9token": (to "844") I can't find you, nor do I know where I am,

- 89 "844": (to "9token") Yes I am very slow and I do not know what I am doing
- 95 "9token": (to "844") I think I am walking to the library, but it never gets any closer
- 213 "sixtet": (to "5star") I have no idea how to navigate

Disorientation can be attributed to three factors: 1) subjects' lack of knowledge about technical features of the environment (e.g., how avatars work, and how to move them around the environment); 2) the relative paucity of signs within the environment to indicate where one is located, or how to find other locations; and 3) technical problems and design inconsistencies within the environment itself. With regard to the third factor, the first trial experienced network lag in the middle of the experiment, causing some subjects' avatars to slow down or cease to respond to commands temporarily. This is the problem referred to by 844 in line 89: "Yes I am very slow". Active Worlds environments also have a disorienting design feature: the background graphics, although photo-realistic (the library building can be seen in the upper left corner of Figure 2), are not part of the navigable graphical environment; one can 'walk' continuously toward them without ever being able to reach them, as noted by 9token in line 95. A further inconsistency is that in addition to the 'real' library in the background, there is a virtual (graphical) library in iUni, which subjects must locate as part of the treasure hunt task (see Appendix). Perhaps for these reasons, expressions of disorientation persist and even tend to increase over time in the chat logs.

Less commonly, subjects also comment explicitly when they start to understand how the technological environment works. Example (6) illustrates expressions of emergent understanding.

- (6) 174 "2cool": (to "1derful") where are you going?
- 175 "2cool": can anyone hear me?
- 176 "1derful": I'm here...I guess I was too far away.

192 "9token": has anyone walked over a blinking blue dot and did it mess
you up

In line 176, 1derful expresses an emergent understanding that when an avatar moves away, it can no longer 'hear' others. In line 192, 9token expresses a suspicion that a feature of the graphical environment, a blinking blue 'dot', causes something undesirable to happen to an avatar that walks over it.^{21[21]}

Example (7) illustrates expressions that reflect mastery and/or full understanding of the virtual environment.

(7) 198 "02" said 'hold CTR and press forward! you can run!!
220 "5star": (to "sixtet") I'm finally getting the hang of it, but I think
the network was my biggest problem.

02's comment in line 198 demonstrates de facto mastery of a navigation command. In line 220, 5star comments that she is 'finally getting the hang of' navigating in the virtual world. Although relatively infrequent, expressions of mastery of the environment increase significantly ($p=.0008$) over time, and emergent understandings also tend to increase.

5.1.4. Task focus

Most messages in the chat logs are either about the task or the virtual environment; relatively few are entirely unrelated to the activity at hand. Of those that are, most are greetings that occur near the beginning of the experiment. Playful behavior is most likely to occur in this initial greeting phase, in the form of whimsical greetings (8) and nonsense words or syllables (9), as subjects test the messaging capability of the system.

(8) 10 "sixtet": hola

14 "3beE": (to "sixtet") pssst
 24 "sixtet": (to "fortune") toodles

(9) 20 "13" said ':rats~'
 21 "09" said 'rats'
 23 "09" said 'ratssssssssssssssss'
 25 "09" said 'bund'
 27 "09" said 'bunrats'
 28 "13" said 'the rat!'
 29 "09" said 'bundrat!'

(9) is the only example of extended interactive play in the data, and it appears near the beginning of the chat log in the second trial. As predicted, off-task messages unrelated to the environment show a significant decrease over time ($p < .0001$).

Most off-task messages are oriented to the virtual world. In addition to the expressions of disorientation in example (5), subjects ask and instruct one another about the navigational features of the environment. This includes means of movement that are not necessarily practical for completing the treasure hunt task, but that are unusual or amusing, such as 'flying' and 'jumping'.

(10) 53 "13" said 'Hey! u can actually jump around ! '
 56 "19" said 'testing...'
 58 "19" said 'wow~~ '
 60 "13" said 'can anyone seee me jumping?'
 62 Harry said 'yes :-)'

Although we had expected it to decrease as users became more familiar with the environment, talk about location/movement in fact increases significantly over time ($p = .0012$). This is consistent with the view that navigation in the virtual world is distracting to inexperienced users,

and suggests further that it becomes more distracting (or that users verbalize their distraction more) over time. In contrast, utterances that reference perception of the virtual environment display a mixed pattern. Talk about 'seeing' decreases ($p=.0048$), consistent with the Adaptation hypothesis, but talk about 'hearing' increases, albeit not significantly. This may reflect a tendency for subjects to rely decreasingly on the visual perception of their avatars, which as we have seen can be confusing, and increasingly on the text in the chat window, which works in a more familiar manner than the graphics window and includes more participants in the conversation. The shift from an orientation towards 'seeing' to 'hearing' (reading) constitutes user adaptation to the system.

Finally, task-focused messages increase significantly over time ($p=.05$), starting with identifying one's team members. The shift to task focus took place roughly 15% of the way through each chat log, in terms of number of messages. The first task-focused message from each trial is shown in example (11).

(11) 34 "2cool": (to 3beE") hey 3, we're on the same team -- where should we
meet
95 "12rp" said 'ok where is 9, 10, and 11 (we are team 3)?'

That task-focused messages increase over time directly supports the Adaptation hypothesis.

5.1.5. Task schema

Talk about the treasure hunt task follows the logical sequence of moves in the hypothesized task schema. That is, subjects first identify their team members (example 11), then set meeting places (example 3), coordinate their team's activities (example 1), move about the environment seeking information, and finally report their answers or their lack of success in the search (example 4).

The succession of moves is represented visually in figure 4, which plots the distribution of

utterances (diamonds) expressing each move (indicated by code numbers along the y-axis). Time is represented along the x-axis as Z scores relative to the mid-points of the two experimental trials, to control for differences in the sample sizes. Note that the lower-numbered moves of the schema cluster in the left-hand corner of the figure, at the beginning of the experiment, and the higher-numbered moves cluster in the right-hand corner, at the end of the experiment, as predicted. Change over time is significant in the expected direction for moves 1 ($p=.034$), 4 ($p=.0001$), and 5 ($p<.0001$), and tends in the predicted direction for moves 2 and 3.

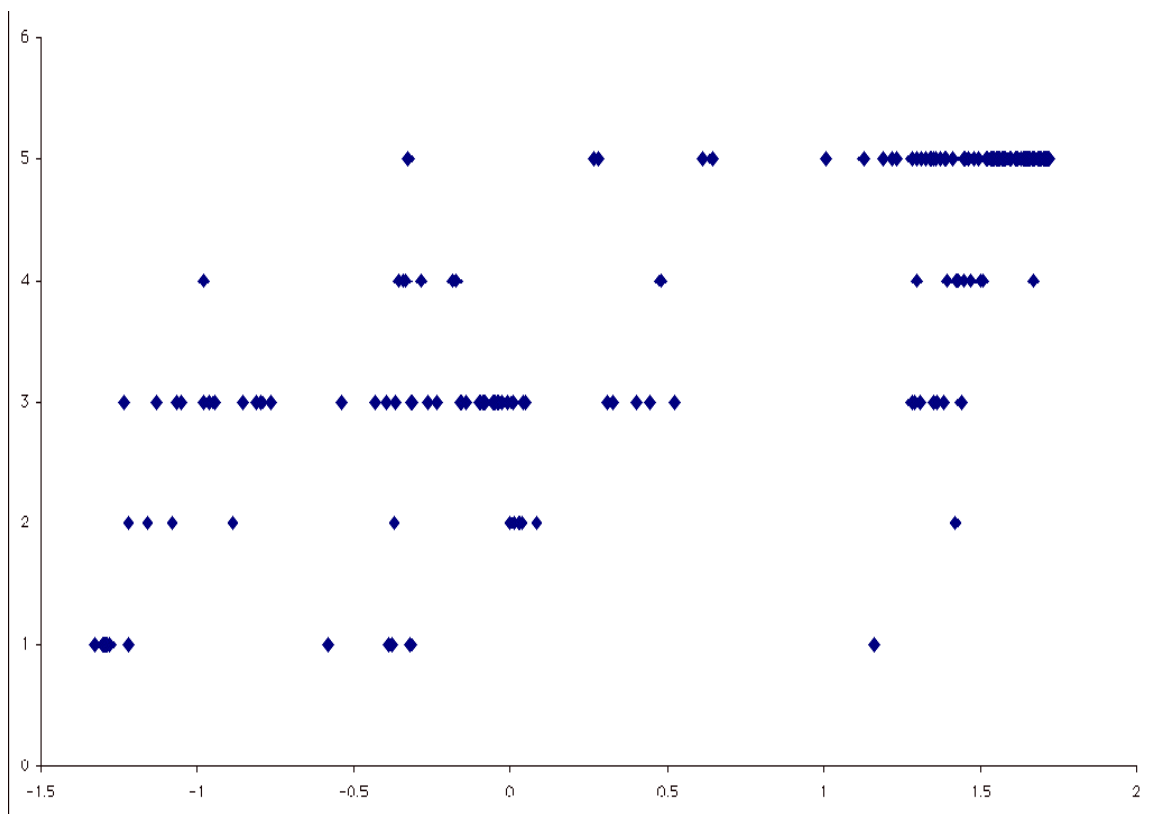


Figure 4. Distribution of the five moves of the task schema over time

This finding indicates that the subjects followed the sequence of moves suggested in the directions for the experiment. Taken together with the low incidence of genuinely off-task communication, it suggests that subjects took the task seriously, and that the disorder apparent in

parts of the log was due to disorientation caused by the virtual world, rather than to playfulness or a disregard for the protocols of the study.

5.2. Differences among teams

As presented above, despite the brief amount of time allotted to subjects to complete the information treasure hunt, their communication patterns altered in the direction of task focus and specificity. We believe that this is what enabled the subjects to complete the task to the extent they were able to do so. (No team found all 10 answers, and several found none.) Further support for this interpretation comes from evidence that the more successful teams were more task focused and specific than the less successful teams, when time is excluded as a factor.

Teams 1-3 are from the first trial, and teams 4-8 from the second trial. The most successful teams, in terms of number of results reported, were Teams 1, 3, 6 and 8. These teams reported finding between three and five answers to the 10 questions in the information treasure hunt. In contrast, Teams 2, 4, 5 and 7 reported finding zero or one answer. Table 1 summarizes the behaviors favored and avoided by the successful teams.^{22[22]}

Table 1. Communicative behaviors favored and disfavored by successful teams

Favored	Disfavored
Fixed spatial reference Addressing team Talk about seeing and hearing Task focus Reporting results	Deixis Addressee unspecified Off-task messages Talk about location/movement

The behaviors favored and avoided by the unsuccessful teams are almost the exact inverse of

those in Table 1. These are summarized in Table 2.

Table 2. Communicative behaviors favored and disfavored by unsuccessful teams

Favored	Disfavored
Deictic verbs Addressee unspecified Off-task messages	Fixed spatial reference Addressing team Reporting results

A generalization that emerges is that behaviors hypothesized to occur later in the task for the sample as a whole (e.g., task focus, specific spatial reference, reporting results) are used more by successful teams, and behaviors hypothesized to occur earlier in the task overall (e.g., off-task messages, deixis, and messages with an unspecified addressee) are used more by unsuccessful teams. This result provides independent verification that the more specific, focused behaviors are more efficacious for accomplishing the collaborative task, and hence adaptive in the given environment.

6. Discussion

6.1. Summary

The results of this study show that when first introduced to the graphical 3-D virtual world, subjects use face-to-face strategies such as deixis and lack of explicit address, which rely on a shared visual context and awareness of the location and orientation of the bodies of one's interlocutors. Such strategies are not typical of text-only chat rooms (cf. Werry 1996); the reason for their use is presumably the pseudo-realistic, navigable graphical interface and the use of humanoid avatars, which create the expectation that space, movement and perspective will function as in the physical world. Accordingly, subjects employ face-to-face discourse strategies.

However, the expectation of realism is repeatedly violated as soon as the subjects try to communicate or navigate in the 3-D virtual world. Unlike in the physical world, participants in conversations in open spaces in 3-D virtual worlds can ‘hear’ farther than they can ‘see’. ‘Seeing’ is further constrained by avatars' limited range of vision, in combination with an inability to turn their eyes or heads. These perceptual limitations make deictic reference and inexplicit address difficult to resolve. The environment itself also behaves in unrealistic and inconsistent ways which, in conjunction with a relative lack of maps or directional signs, make purposeful navigation difficult.^{23[23]} Taken together, these seemingly realistic *but in fact unrealistic* properties of the system make face-to-face reference strategies impracticable.

The information treasure hunt task required subjects to navigate, and communicate, purposefully. In order to accomplish this, the subjects struggled but ultimately adapted their linguistic behaviors. Out of 22 behaviors analyzed, half showed significant change over time; of those changes, 82% were in the predicted direction of greater specificity, context independence, and task focus. Moreover, most of the non-significant results exhibited trends in the predicted direction. The only two results that significantly contradict the hypotheses are interrelated. Talk about location and movement increased, as did deixis, which was often used in talk about location and movement (e.g., 'here', 'go'). The subjects were clearly preoccupied with navigation, which continued to cause them difficulties through the end of the task, even as they mastered some skills and succeeded in collecting some of the information required for the treasure hunt. The results of the comparison of successful and unsuccessful teams lends further support to the notion that specificity and focus are strategies that were adaptive to task at hand.

6.2. Implications

The above findings point to challenges that users of 3-D virtual worlds for structured collaborative purposes are likely to encounter, and suggest successful (and unsuccessful) strategies for responding to those challenges. As such, the findings have both theoretical and

practical implications.

6.2.1. Media richness theory

In the present study, novice users transferred face-to-face discourse strategies to a 3-D virtual world. This is consistent with a fundamental premise underlying media richness theory, as articulated by Durlak (1987; quoted in Steuer 1996):

First-person experiences of the real world represent a standard to which all mediated experiences are compared, either mindfully or otherwise; face-to-face interaction with other humans is used as a model for all interactive communication.

However, face-to-face strategies worked poorly in iUni because of ways in which the virtual world differs from the unmediated, physical world. Indeed, the illusion of realism created by a virtual world can mislead users into making inappropriate communicative choices. Similar observations have been made in other 3-D virtual world studies by Hindmarsh et al. (2000, for deixis) and Darken (1995, for navigation), and in studies of video-mediated communication by Barnard et al. (1996, for deixis) and Heath and Luff (1991, for turn-taking). This evidence challenges the view that richer media necessarily facilitate more natural (i.e., more face-to-face-like) communication.

It might be objected that the problem is not with media richness theory, but rather with the limited knowledge and resources presently available to create virtual worlds. If we could only “do realism” better, the argument goes, the interface would become transparent, and interaction could transpire as effortlessly as if the interlocutors were face-to-face. According to this view, the misleading properties of 3-D virtual worlds such as iUni should be modified—and the design of 3-D virtual worlds in general should proceed—in the direction of increased realism (see, e.g., Damer, Kekenés and Hoffman 1996; Vilhjálmsón 1997).^{24[24]}

An alternative interpretation calls into question the fundamental assumption of media richness theory that face-to-face communication should be privileged as the model for understanding collaborative media. According to this view, face-to-face communication is just another medium—that of spatial and temporal proximity—which like other media has its strengths and limitations (Hollan and Stornetta 1992).^{25[25]} If face-to-face communication is not privileged, it follows that rich media (i.e., those that seek to imitate face-to-face communication) are not automatically superior to leaner media. Applied to virtual worlds, these criticisms suggest that complex graphics and multidimensional navigation do not necessarily facilitate—and under some circumstances, may actually detract from—users' communicative goals.

6.2.2. 3-D virtual world design

We believe that face-to-face discourse should not be the unquestioned standard against which computer-mediated discourse is measured, nor should imitating face-to-face communication be the primary goal of 3-D virtual world design. As Hollan and Stornetta (1992) point out, an imitation, by definition, can never be as good as the original. Rather, we should seek to enhance the usability of virtual worlds so that they can better meet the communicative requirements of the tasks for which they are intended. This might involve minimizing distractions, optimizing strengths, and adding features to compensate for the virtual system's inherent weaknesses, even if the resulting designs lack realism.

Various recommendations have been made to add features to compensate for weaknesses in current 3-D graphical systems. Damer et al. (1996) recommend increasing avatar capabilities and standardizing the navigational cues in the interface to enhance the usability of the system. Hindmarsh et al. (2000) propose enhancing social interaction through visually indicating an avatar's range of vision so that others know what it is seeing. In addition, the findings of the present investigation suggest a need for expanding the avatar's visual range, and its ability to visually scan the environment without having to reorient its body. Work is being undertaken to

build "peripheral vision" into virtual worlds (Robertson, Czerwinski and van Dantzich 1997), in an attempt to address this need. In the absence of a wide visual range, the whisper command (that allows users to communicate even when they can not see or hear one another) proved useful in the present study. Along similar lines, recent work by Börner, Hazlewood and Lin (2002) aims to extend the Active Worlds browser by adding advanced spatial and social navigation guides such as overview maps with markers for user positions, as well as visualizations of user trails and interactions (chat, web access, teleport usage, etc.).

The last two proposals depart from realism in an attempt to facilitate social navigation. Removing some features intended to be realistic, but which the present study has shown to be disorienting or distracting, could further enhance task-based collaboration. An example of such a feature is the panoramic backdrop in many Active Worlds. More generally, the requirement to manipulate an avatar and attend to other avatars imposes a considerable cognitive burden on the user, competing with (and potentially distracting from) the activities of communication and navigation. Graphical embodiments, which are directly premised on an attempt to imitate face-to-face communication, might beneficially be eliminated from non-recreational virtual worlds, and their communicational, navigational and deictic functions replaced by other features.^{26[26]} (A similar proposal has been made with regard to head shots in object-focused video conferencing; see Hinds 1998, cited in Walther 1999; Whittaker 1995; Whittaker and O'Conaill 1993). The resulting interfaces would not necessarily be less complex, but the complexity would be conditioned by principles other than the need to imitate face-to-face communication, such as the requirements of the tasks for which the systems are designed.

7. Conclusions

The results of this study are subject to several qualifications. The discourse analyzed was produced under quasi-experimental conditions, rather than for the subjects' own communicative purposes. As in much experimental research, the task (an 'information treasure hunt') is not

typical of casual interaction in most online environments. Thus caution should be exercised in generalizing the findings to other environments and other purposes of communication. In particular, they might not apply to recreational, unstructured uses of 3-D virtual worlds.

At the same time, due to the exploratory nature of the investigation, individual variables that might condition communicative behaviors were not controlled, as they would have been in a strict experimental design. Thus it is not possible to determine that a particular structural feature of the virtual environment had a greater effect on user behavior than another feature. Future experimental studies should vary features of the environment, the task, and characteristics of user groups systematically in order to isolate causal factors. For example, the subjects had only one hour in which to complete a challenging task in an unfamiliar environment. It is possible that more experienced users of iUni, or novice users given more time and/or advance training, would have been less disoriented, and adopted different linguistic strategies as a consequence.

Nonetheless, it is significant that novice users when encountering the 3-D graphical environment for the first time opted to employ face-to-face reference strategies, rather than text-based forms of reference with which they were familiar from other modes of synchronous computer-mediated communication. This suggests that the pseudo-realism of the environment, despite some obviously artificial aspects, led them to over-attribute properties of the real world to the environment. The speed with which the novice users shifted to a more text-based, decontextualized strategy is also significant, suggesting that the misleading effects of the graphical world can be overcome to some extent through experience (and the use of leaner media such as text).

Ultimately, our goal is to understand multimedia systems in their own terms, rather than as alternative forms of face-to-face communication. New systems sometimes require new methods of analysis. A future direction that we have already begun to explore is to analyze avatar movement and location data in conjunction with textual logs. Capturing the dynamic sequences of actions taking place in the graphical window, e.g., by using video recording software, would

enable more precise analysis of the visual conditions under which certain discourse behaviors occur. As a step towards this goal, we have begun collecting numerical data about the position of avatars when they ‘speak’, and using such data to generate visualizations of subjects' navigational trajectories in relation to their talk.

Appendix: iUni Treasure Hunt Questions

1. How many student residences are there?
2. Write down the names of all students who have residences.
3. Find 'Magster's Mansion' and answer: How many worlds are linked from her menu system?
4. In the 'Meeting Place' area you will find a sign that links to lecture slides. What was the topic of this particular lecture?
5. What are the four main resources that are available in the 'Library'?
6. How many different areas are in the 'Art Café'?
7. In the 'Quest Atlantis' area: Is 'Jonassan' displayed in the 'Tower of Power'?
8. How many household appliances are linked from the 'Science House'?
9. In the 'Natural Disaster' area, how many different types of buildings are displayed?
10. How many video streams are accessible in the 'Virtual Collaboration' area?

References

- Allwood, Jens and Ralph Schroeder (2000): Intercultural communication in a virtual environment. *Intercultural Communication*, vol. 4.
<<http://www.immi.se/intercultural/>>
- Barnard, Philip, Jon May and Daniel Salber (1996): Deixis and points of view in media spaces: An empirical gesture. *Behavior and Information Technology*, vol. 15, no. 1, pp. 37-50.
- Baron, Naomi (1998): Letters by phone or speech by other means: The linguistics of email. *Language and Communication*, vol. 18, pp. 133-170.
- Baym, Nancy (1996): Agreements and disagreements in a computer-mediated discussion. *Research on Language and Social Interaction*, vol. 29, pp. 315-346.
- Benford, Steve, John Bowers, Lennart Fahlen, John Mariani and Tom Rodden (1994): Supporting co-operative work in virtual environments. *The Computer Journal*, vol. 37, no. 8.
- Börner, Katy (2001): Using Active Worlds technology to build an iUniverse of 3-D collaborative learning environments. *IEEE Learning Technology newsletter*.
<http://lttf.ieee.org/learn_tech/issues/january2001/>
- Börner, Katy, Richard Hazlewood and Sy-Miaw Lin (2002): Visualizing the spatial and temporal distribution of user interaction data collected in three-dimensional virtual worlds. *Proceedings of the Sixth International Conference on Information Visualization, London, England, July 10-12, 2002*. IEEE Press, pp. 25-31.
- Cherny, Lynn (1999): *Conversation and Community: Chat in a Virtual World*. Stanford, CA: CSLI Publications.

- Cho, Natasha (In press): Linguistic features of electronic mail in the workplace: A comparison with memoranda. In S.C. Herring (ed.): *Computer-Mediated Conversation*. Cresskill, NJ: Hampton Press.
- Croon Fors, Anna and Mikael Jakobsson (2000): Beyond use and design—the dialectics of being in virtual worlds. Paper presented at *Internet Research 1.0: The State of the Interdiscipline*, Lawrence, Kansas.
- Csikszentmihalyi, Mihaly (1990): *Flow: The Psychology of Optimal Experience*. New York: Harper and Row.
- Daft, Richard and Robert Lengel (1984): Information richness: A new approach to managerial behavior and organization design." *Research in Organizational Behavior*, vol. 6, pp. 191-233.
- Daft, Richard and Robert Lengel (1986): Organizational information requirements, media richness and structural design. *Management Science*, vol. 32, no. 5, pp. 554 - 571.
- Damer, Bruce, Christina Kekenes and Terrel Hoffman (1996): Inhabited digital spaces. *ACM CHI '96 Companion*, pp. 9-10.
- Darken, Rudy P. (1995): *Wayfinding in Large-Scale Virtual Worlds*. Doctoral dissertation, The George Washington University, Washington, D.C.
- Darken, Rudy P. and John Sibert (1996): Wayfinding strategies and behaviors in large virtual worlds. *Proceedings of CHI 1996*.
<http://www.acm.org/sigchi/chi96/proceedings/papers/Darken/Rpd_txt.htm>
- Eep2. (2000): *Active Worlds History*. <<http://tnic.com/eep/aw/history.html>>
- Fillmore, Charles J. (1997): *Lectures on Deixis*. Stanford, California: CSLI Publications.
- Fillmore, Charles J. (1982): Towards a descriptive framework for spatial deixis. In R. Jarvella & W. Klein (eds.): *Speech, Place and Action: Studies in deixis and related topics*. New York: John Wiley, pp. 31-59.
- Fischer, Kerstin and Thora Tenbrink (2002): Video conferencing in a transregional research cooperation: Turn-taking in a new medium. *Internationale Fachtagung*

- "Videokonferenz: Grundlagen - Anwendungen - Perspektiven, Universität Essen, 6-8. November 2002.
- Grayson, David and Lynne Coventry (1998): The effects of visual proxemic information in video mediated communication. *ACM SIGCHI Bulletin*, vol. 30, no. 3, pp. 30-39.
- Hård af Segerstad, Ylva (2002): *Use and Adaptation of Written Language to the Conditions of Computer-Mediated Communication*. Ph.D. Dissertation, Department of Linguistics, Göteborg University, Sweden.
- Harrison, Steve and Paul Dourish (1996): Re-place-ing space: The roles of place and space in collaborative systems. *Proceedings of the Conference on Computer Supported Cooperative Work..* ACM Press, pp. 67-76.
- Herring, Susan C. (1996): Two variants of an electronic message schema. In S.C. Herring (ed.): *Computer-Mediated Communication: Linguistic, Social and Cross-Cultural Perspectives*. Amsterdam: John Benjamins, pp. 81-108.
- Herring, Susan C. (2001): Computer-mediated discourse. In *The Handbook of Discourse Analysis*, D. Schiffrin, D. Tannen and H. Hamilton (eds). Oxford: Blackwell Publishers, pp. 612-634.
- Herring, Susan C. (2002): Computer-mediated communication on the Internet. *Annual Review of Information Science and Technology*, vol. 36, pp. 109-168.
- Herring, Susan C. (2003): Computer-mediated discourse analysis: An approach to researching online behavior. *Designing for Virtual Communities in the Service of Learning*, S. A. Barab, R. Kling and J. H. Gray (eds.). New York: Cambridge University Press.
- Hindmarsh, Jon, Mike Fraser, Christian Heath, Steve Benford and Chris Greenhalgh (2000): Object-focused interaction in collaborative virtual environments. *ACM Transactions on Computer-Human Interaction*, vol. 7, no. 4, pp. 477-509.

- Hollan, Jim and Scott Stornetta (1992): Beyond being there. *Proceedings of CHI 1992*.
New York: ACM Press, pp.119-125.
- Kiesler, Sara, Jane Siegel and Timothy McGuire (1984): Social psychological aspects of computer-mediated communication. *American Psychologist*, vol. 39, pp. 1123-34.
- Kolko, Beth (1999): Representing bodies in virtual space: The rhetoric of avatar design. *The Information Society*, vol. 15, pp. 177-186.
- Krikorian, Dean, Jae-Shin Lee, T. Makana Chock and Chad Harms (2000): Isn't that spatial? Distance and communication in a 2-D virtual environment. *Journal of Computer-Mediated Communication*, vol. 5, no. 4.
<<http://www.ascusc.org/jcmc/vol5/issue4/krikorian.html>>
- Levinson, Stephen (1983): *Pragmatics*. Cambridge: Cambridge University Press. Chapter 2.
- Lombard, Matthew and Theresa Ditton (1997): At the heart of it all: The concept of presence. *Journal of Computer Mediated Communication*, vol. 3, no. 2.
<<http://www.ascusc.org/jcmc/vol3/issue2/lombard.html>>
- Lyndon, Donlyn and Charles W. Moore (1996): *Chambers for a Memory Palace*.
Cambridge: MIT Press.
- Lyons, John (1977): *Semantics*. Cambridge: Cambridge University Press. Vol. 2, Chapter 15.
- Manninen, Tony and Jani Pirkola (1997): Comparative classification of multi-user virtual worlds. <http://www.tol.oulu.fi/~tmannine/game_design/multi-user_virtual_worlds.pdf>
- Manovich, Lev (1996): Virtual worlds. *Telepolis*.
<<http://www.heise.de/tp/english/special/sam/6030/1.html>>
- McClellan, J. (1996): Alpha World: Police pages.
<<http://binky.paragon.co.uk/it/issue19/Alpha/Alpha.html>>

- McDonough, Jerome P. (1999): Designer selves: Construction of technologically mediated identity within graphical, multiuser virtual environments. *Journal of the American Society for Information Science*, vol. 50, pp. 855-869.
- McMahon, Tamara M. and Katy Börner (Under consideration): A study of the effects of web page panel size and layout density on information access in a 3-dimensional collaborative virtual world. Special issue of *Journal of Digital Libraries: Usability for Digital Libraries*, A. Blandford and G. Buchanan (eds).
- Mitchell, Don (1995): From MUDs to virtual worlds.
<<http://research.microsoft.com/scg/papers/3DVW.htm>>
- Murray, Denise (1988): The context of oral and written language: A framework of mode and medium switching. *Language in Society*, vol. 17, pp. 351-373.
- Naper, Ida (2001): System features of an inhabited 3D virtual environment supporting multimodality in communication. *Proceedings of the 34th Hawaii International Conference on System Sciences*. Los Alamitos, CA: IEEE Computer Society Press.
- Naper, Ida (In press): Conversation in a multimodal 3D virtual environment. In S. Herring (ed.): *Computer-Mediated Conversation*. Cresskill, NJ: Hampton Press.
- Neal, Lisa (1997): Virtual classrooms and communities.
<<http://www3.ncsu.edu/dox/NBE/neal/nealtitle.htm>>
- O'Malley, Claire, Steve Langton, Anne Anderson, Gwyneth Doherty-Sneddon and Vicki Bruce (1996): Comparison of face-to-face and video-mediated interaction. *Interacting with Computers*, vol. 8, no. 2, pp. 177-192.
- Paolillo, John C. (2001): Language variation in the virtual speech community: A social network approach to Internet Relay Chat. *Journal of Sociolinguistics*, vol. 5, no. 2, pp. 180-213.
- Robertson, George G., Stuart K. Card and Jock D. Mackinlay (1993): Nonimmersive virtual reality. *Computer*, vol. 26, pp. 81-83.

- Robertson, George G., Mary Czerwinski and Maartin van Dantzich (1997): Immersion in desktop virtual reality. *Proceedings of the 10th annual ACM symposium on User Interface Software and Technology*, pp. 11-19.
- Scannell, Beth (1999): *Life on the border: Cyberspace and the frontier in historical perspective*. <<http://www.kether.com/thesis/context.html>>
- Sellen, Abigail (1992): Speech patterns in video-mediated conversations. *Proceedings CHI'92*, pp. 49-59.
- Sellen, Abigail (1995): Remote conversations: The effects of mediating talk with technology. *Human-Computer Interaction*, vol. 10, pp. 401-444.
- Short, John, Ederyn Williams and Bruce Christie (1976): *The Social Psychology of Telecommunications*. London: John Wiley.
- Steinhardt, Steve (n.d.): An interview with Dave Gobel, chairman and founder of Worlds Inc. *Online Magazine*. <<http://www.online-magazine.com/gobel1.htm>>
- Steuer, Jonathan (1992): Defining virtual reality: Dimensions determining telepresence. *Journal of Communication*, vol. 42, no. 4, pp. 73- 93.
- Suler, John (1996): *The Psychology of Cyberspace*.
<<http://www.rider.edu/users/suler/psycyber/psycyber.html>>
- Vilhjálmsón, Hannes (1997): *Autonomous communicative behavior in avatars*. Master's thesis, MIT Media Lab. <<http://web.media.mit.edu/~hannes/msthesis/>>
- Walther, Joseph (1996): Computer-mediated communication: Impersonal, interpersonal and hyperpersonal interaction. *Communication Research*, vol. 23, no. 1, pp. 3-43.
- Walther, Joseph (1999): Visual cues and computer-mediated communication: Don't look before you leap. <<http://www.renselaer.edu/~walthj/ica99.html>>
- Werry, Christopher (1996): Linguistic and interactional features of Internet Relay Chat. In S. Herring (ed.): *Computer-mediated communication: Linguistic, social and cross-cultural perspectives*. Amsterdam: John Benjamins, pp. 47-63.

Whittaker, Steve (1995): Rethinking video as a technology for interpersonal communications: Theory and design implications. *International Journal of Man-Machine Studies*, vol. 42, pp. 501-529.

Whittaker, Steve and Brid O'Conaill (1993): Evaluating videoconferencing. In *Proceedings of CHI'93 Human Factors in Computing Systems*. New York: ACM Press.

Yates, Simeon and David Graddol (1996, July): "I read this chat is heavy:" The discursive construction of identity in CMC. Paper presented at the 5th International Pragmatics Conference, Mexico City.

Notes

^{1[1]} Active Worlds <<http://www.activeworlds.com/>>

^{2[2]} The tendency towards realism in virtual worlds is not universal. Steinhardt (n.d.) notes that the Japanese prefer fantasy graphics.

^{3[3]} For overviews of language-focused (text-based) CMC research, see Herring (2001, 2002, 2003).

^{4[4]} An example of a video-mediated communication system is CUseeMe, a video plus audio plus text interface that allows two or more participants to communicate in real time over the Internet. See Herring (2002) for a summary of the history of video-mediated desktop environments.

^{5[5]} See Herring (2002) for a summary of the history and use of desktop graphical virtual reality environments.

^{6[6]} The reason for this is the availability of the 'emote' command in MUDs, and "the strong and rich language support" available in text-based virtual worlds more generally (Manninen and Pirkola 1997).

^{7[7]} WorldsAway is "2 1/2 dimensional", meaning that one object, or avatar, may move in front of or behind another (Scannell 1999).

^{8[8]} "Designing a new car with virtual reality." SGI

<<http://www.sgi.com/features/2001/may/daimlerchrysler/>>

^{9[9]} Dead reckoning navigation is a method of reckoning or reasoning one's position relative to something stationary, based on the direction of movement and the estimated speed since the last known position.

^{10[10]} The reason avatars in 3-D virtual worlds typically perceive a narrow range of vision is that distortion results when wider angles are used, compromising the illusion of realism.

^{11[11]} This need not be the case—some graphical worlds have avatars that represent animals, such as birds (Naper 2001), or inanimate objects, such as billiard balls or automobiles (Suler 1996).

^{12[12]} Regular users can become 'citizens' of the world, in which case they may choose from a range of avatars, including a bird. 'Tourists', or casual visitors, must select either the male or the female default avatar. (See Naper In press for discussion of the differences between the functionality accorded to 'tourists' and 'citizens' in the Patagonia Active World.) All of the subjects in the present study logged into the system as 'tourists'.

^{13[13]} The third author was involved in the design of iUni as a student in the course. She was responsible for designing the 'Natural Disasters' area.

^{14[14]} Bots—machine programs—have a "hearing" range of 200 meters.

^{15[15]} When no whisper command is available, if team members are more than 120 meters apart, there is almost no way for them to find each other except by moving from one corner of the world to the next (what Darken 1995 calls the 'naive search pattern') and calling out, hoping they will respond. Perhaps for this reason, most conversations in our study took place at or near the default entrance point to the site, the "ground zero", a location where other avatars can most reliably be found.

^{16[16]} Because any given avatar has a limited "hearing" (reading) range, no one participant can log messages "spoken" in all parts of the environment. To get around this problem, we instructed each subject to save a log from their avatar's perspective, and we later

combined the individual logs into a single transcript for each experimental trial, eliminating redundant messages. This task was facilitated by the availability of time stamps in the second trial that indicated the hour, minute, and second each message was posted.

^{17[17]} The 'whisper' option was not functioning during the second treasure hunt, thus this coding category is only analyzed for the first set of results.

^{18[18]} We did not distinguish between correct and incorrect answers in this calculation, as in keeping with the difficulty of the task, most of the answers reported were incorrect. We considered groups that were able to navigate to the locations specified in the treasure hunt instructions to have been more successful than groups that were not able to find those locations, even if the information they reported about them was inaccurate. We counted answers reported in the chat logs, rather than information written down on paper, since subjects (especially in the second trial) could have "overheard" other teams reporting their findings and simply written them down. In contrast, there was not much repetition of treasure hunt findings across teams in the chat logs themselves.

^{19[19]} In support of this interpretation, another subject used (lower-case) 'k' in the same sense a few lines later in the chat log, and it did not cause confusion.

^{20[20]} Werry (1996) terms the practice of including the addressee's name in a chat message 'addressivity'.

^{21[21]} The blue "dots" are teleport sites associated with a 'virtual tour' that one can take of iUni. When an avatar steps on one, it is teleported to the next location on the tour route.

^{22[22]} All results listed in Tables 1 and 2 are statistically significant ($p \leq .05$); that is, the teams in one category differ significantly from the teams in the other category with respect to the listed behaviors.

^{23[23]} Cf. Darken and Sibert (1996): "The lack of directional cues and spatial organization in the control treatment led to ineffective search strategies and frequent disorientation."

^{24[24]} Vilhjálmsón's (1997) efforts to create life-like avatars with animated facial features and head movements that accompany speech illustrate this approach.

^{25[25]} Face-to-face communication and rich media such as video conferencing can be too rich for some purposes (Walther 1999). Walther (1996) argues that text-only

asynchronous CMC—a "lean" medium according to media richness theory—can lead to better task outcomes and more positive ('hyperpersonal') social relationships, given sufficient time.

^{26[26]} Avatar communication functions could be replaced by features that allow users to identify and chat with others regardless of what part of the environment they are viewing, such as maps, attentional awareness indicators, whispering and text chat. Navigation could proceed efficiently by calling up different locations within the environment, rather than by walking or flying through the environment. In order to interpret deictic reference, users could access one another's computer screens directly (e.g., in a video window in a corner of their own screens), effectively enabling them to see through one another's eyes. Where the goal is object-focused collaboration, an interface of this sort could facilitate deictic reference (paradoxically, by making it *less* like face-to-face reference) and enhance usability by leaving users more attentional resources to focus on the task at hand. It would also free designers of virtual environment from constraining metaphors of physical space (landscapes, buildings, paths, gravity, etc.) to explore new and possibly better means for representing and organizing information (e.g., 'memory palaces', Lyndon and Moore 1996).