Talking and Building:
Two Studies of Collaboration in Second Life

Greg Wadley
visiting researcher at Palo Alto Research Center (PARC)
3333 Coyote Hill Road, Palo Alto, CA
gwadley@parc.com

INTRODUCTION

This position paper discusses two studies of user communication and collaboration in Second Life (SL). One looked at the impact of voice communication after its introduction to SL in 2007. The other examined collaborative building in SL. Voice is considered by many to be essential for workplace use of VWs; however its introduction was controversial in the SL user community. The ability to build 3d content collaboratively sets SL apart from most other virtual worlds, and allows us to update our understanding of collaboration in 3d. Here I describe these studies and discuss the results with regard to the use of virtual worlds in the workplace.

STUDY 1: VOICE COMMUNICATION IN SECOND LIFE

In 2007/8 I conducted a study into the effects of voice on people's experience of SL. This followed previous studies of voice in MMOGs [14] and the Xbox Live network [5].

Voice was introduced to SL in 2007 amidst controversy. Many SL users prefer to communicate by typing text. SL users are often categorized as "immersionists" or "augmentationists". The former seek to maintain an online identity which is independent of their RL identity, a practice that has been documented in older online technologies [13]. The latter however see virtual worlds as communication media and project their real-life identity into SL to interact with other "real" people. It has been argued that immersionists should prefer the relative anonymity of text, while augmentationists should embrace the efficiency and immediacy of voice (eg [1] pp 112-6).

While these categories are popular with SL users, it has been suggested that they be more usefully be thought of as ideal types marking the ends of a spectrum along which real users might be placed [1]. I argue moreover that they represent kinds of use rather than kinds of users.

Text and voice differ along many dimensions that make them more or less suitable to different communication scenarios. I synthesized prior research into media effects (eg [4, 9]) into eight dimensions that may influence a user's choice of medium, and the communication subsequently conducted. To study attitudes to voice I interviewed a wide range of users, from business-people and teachers to artists and role-players. I used this data to identify immersionist and augmentationist preferences along these dimensions. This analysis is illustrated in figure 1.

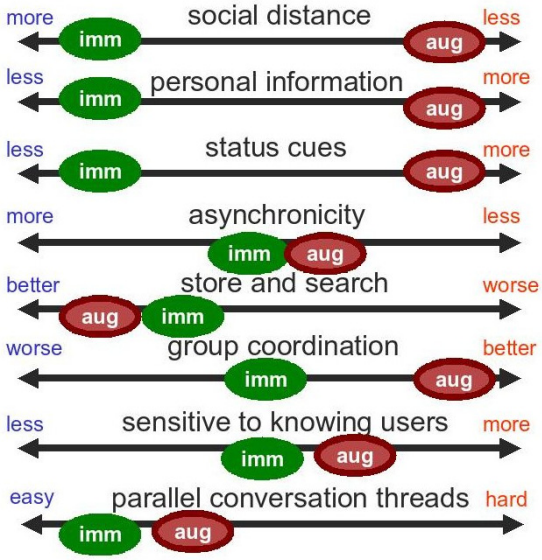

Figure 1: Preferences of SL users for Text (left) and Voice (right)

Immersionists should prefer a communication medium that maximizes the social distance between themselves and other users and which obscures rather than transmits real-life status cues. Augmentationists on the other hand should prefer media that induce the feeling of being "with" their communication partners.

I found that, as predicted, people using SL for teaching or workplace meetings usually preferred voice. However the text channel's better support for message storing and parallel threads was valued.

In practice, SL discussions are often held in text for mundane reasons. VoIP systems in VWs are often less reliable than in-world text chat. In a large enough group someone will usually lack voice hardware or have it incorrectly configured. If one user cannot use voice, many groups revert to text lest that user be completely excluded from discussion. The best communication solution is often a mix of several media [15] and it is interesting to note that newer VWs such as Wonderland emphasize this feature.

STUDY 2: COLLABORATIVE BUILDING IN SECOND LIFE

At PARC, Nic Ducheneaut and I examined collaboration around objects in 3d. Collaboration in virtual-reality has been studied in labs for over a decade (see for example [3]). However while VR has not achieved widespread use,

desktop virtual worlds are popular. The emergence of a VW that allows users to build together offers an opportunity to update our knowledge about collaboration in 3D spaces.

We asked groups of participants to collaborate on building tasks. We recorded their screen video and conversation for later analysis. After the tasks we discussed with participants their building experience, problems faced and solved, and how well the UI supported group building.

Second Life's movable camera

Prior research such as [6] found that the main problem for collaborating groups in VR/CVE was achieving common reference to objects, since a collaborator's vista could not be deduced if their avatar was not visible. However SL's movable camera may make this problem intractable.

A significant difference between the user interface of SL and many other VWs is that SL allows users to move their viewing camera independently of the position and orientation of their avatar. This allows a user to gain multiple perspectives of an object more quickly than is possible by walking an avatar around it. It is in common use for object-related activity such as building, or looking at someone else's creations. Unlike most videogames, which support limited or no camera movement, the SL camera can be moved over a wide area, oriented in any direction including up and down, and zoomed a long way in and out.

At a given moment a user is either looking from their avatar position or their decoupled camera. We call these two modes "in-avatar" and "in-camera". The latter term reminds us that while avatars are visible to other users, a user's camera location is private. The "in-camera" mode is illustrated in Figure 2.

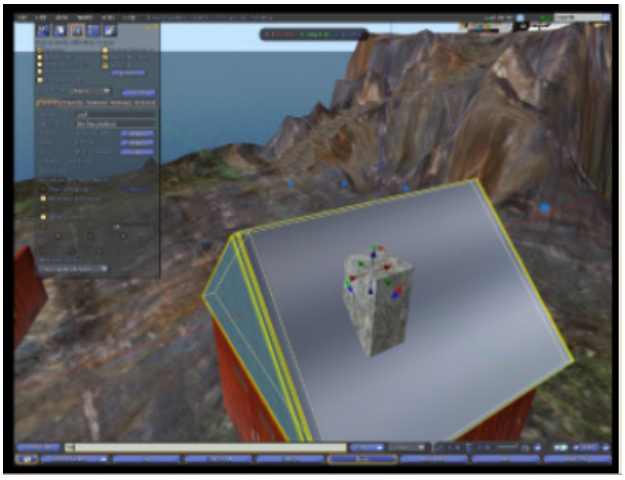

Figure 2: This user is "in-camera" and editing part of a roof. The user's avatar is out of view, on the ground.

Methods

Ten groups of two or three users comprising a total of 22 participants completed two building tasks. Their ages ranged between 20 and 50. Some knew each other before undertaking the task; others were strangers. We chose participants with a range of expertise.

Their first, "house" task was to assemble a building out of existing parts. Each group member was shown a screen-shot of how the house should look when complete. This was similar to the "jigsaw puzzle" task of [7], though in 3D.

For the second, "garden" task, one group member was made leader and shown a picture of the house surrounded by extra objects such as garden furniture and a fence. Only the leader saw the screen-shot, but they could not use the editing tools, instead directing the other participants to build the scene. This was similar to the "helper-worker" arrangement of [7]. The "house" task allowed observation of naturalistic small-group collaboration, while the "garden" task forced more verbal reference to objects.

After each session we conducted focus-group discussions with the participants. We backed up our research by discussing UI problems and strategies with expert SL users discovered in-world and in forums.

Camera independence from avatar

We measured participants' use of the movable camera. On average they spent 43% of task time in-avatar and 57% in-camera. However we noticed patterns of camera use, and to measure these classified our participants into three groups according to expertise. Group A had significant experience while group B were competent users and group C were novices.

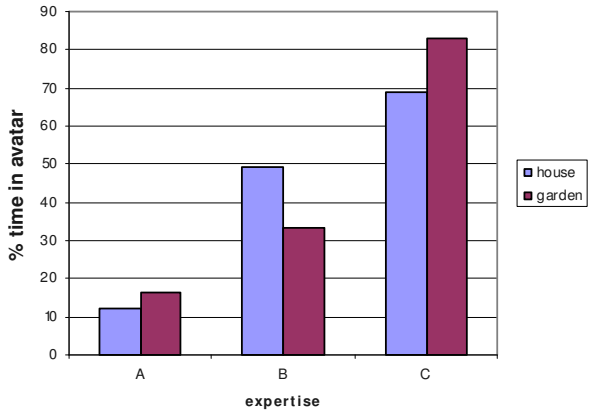

Figure 3: Time spent "in avatar".

Figure 3 illustrates the relationship between expertise and use of camera independence. ANOVA showed that experienced SL users were more inclined to position their camera away from their avatar. Task was not a factor.

When we asked experienced SL users "how would you describe your location right now?", they usually privileged avatar location over camera location. However some participants said that, while building, they felt their avatar was irrelevant or even got in the way.

Techniques for referencing objects

We counted spoken references to objects and places and categorized them, following [8], as "deictic" (expressions relative to an avatar's frame of reference, such as "in front of you"), "intrinsic" (relative to an object's frame of reference, such as "on top of the house"), and "extrinsic" (relative to an external frame of reference, such as compass directions). Although both Cartesian and cardinal frames are available in the Second Life UI, there were few incidents of extrinsic reference.

We added a category: "reference by Name or Property" to count utterances such as "the brown rectangle". ANOVA showed that across groups, reference by name or property was used most, followed by deictic and intrinsic reference. This is illustrated in Figure 4.

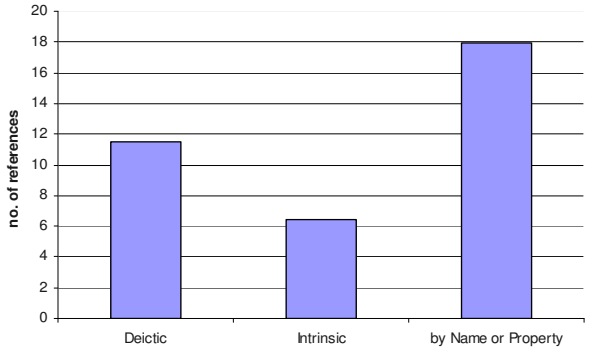

Figure 4: How users referenced objects.

Pointing was rarely used to communicate, and other gestures never used. Some users told us that it was common practice in SL to create prims to mark a location (to be deleted after use), however no-one did this in our trials.

Experienced SL users did not express interest in changing the building UI, though a consistent complaint was that the ownership permission system for objects made group projects harder than they should be. For example objects created by different individuals could not be linked.

The out-of-avatar experience

The ability to easily move the view-camera independently of avatar position represents a major difference between SL and other virtual worlds. At first glance it might impact collaboration in several ways:

1. it is impossible to deduce what a user is looking at;
2. therefore deictic references may be hard to resolve;
3. it is impossible to deduce what objects a user is working with;
4. users have both public (avatar) and private (camera) locations, making group awareness harder;
5. private locations may foster deception and a reduced sense of embodiment in one's avatar.

Prior research often assumed that a user wishing to refer to an object would first deduce what a collaborator could see, from their avatar's location and gaze direction. Users were frequently not successful at doing this [6]. Yet our participants did not have major problems collaborating on an (albeit simple) construction task. Their conversations about objects proceeded reasonably smoothly without becoming mired in problems of reference. We rarely saw a user misinterpret to which object a collaborator referred.

Even experts used deixis relative to a collaborator's avatar, despite knowing the collaborator's camera might be elsewhere. This indicates that deixis is successful often enough, or easy to work around when it fails. It seems that experts are able to maintain a sense of where their avatar is facing when their camera is looking elsewhere. When they could not dereference deixis they temporarily returned to their avatar to translate it. The cost was only the time taken to return to the camera position. This could be easily addressed by adding a "toggle" to switch between the two perspectives.

Avatar proxemics

One might expect that a movable and invisible camera should diminish the relevance of avatar location [10]. But our experience accords with the finding that SL users often position their avatars according to real-world proxemic norms [16]. The avatars of users engaged in communication usually stand face to face even if the users are in-camera. On the other hand, when building, our participants were apt to position their avatars near relevant objects or even forget about them altogether. It seems that SL users see the avatar as a mediator of social interaction, to be "parked" somewhere while building.

It could be argued that to maintain proxemics with one's public embodiment (avatar) while one's private location of focus (camera) is elsewhere, is a form of deception or perception-management such as described by [2].

Prior VW experience and the SL building UI

None of our participants had difficulty navigating their avatar around virtual space, even though not all were frequent users of VWs or games. However many found it hard to use the SL building tools. Translating and rotating an object along orthogonal axes was frustrating for novices. Previous experience with other 3D editing tools substituted for experience with the SL building tools. However experience with games did not, and may have made disembodied camera movement feel unnatural to some.

Some novices seemed to assume that objects were subject to collision detection and gravity, even though this was clearly not so. Experts made use of the fact that objects could hang in the air, pass through other objects, and had a natural orientation along the world's axes, to minimize translations and rotations.

Collaboration and its articulation

Prior research has focused on closely-coupled collaboration, yet we rarely observed this in our study. Participants devised organizational structures that precluded the need for fine-grained collaboration, possibly because the articulation work required for close collaboration represented too high a load. A striking proportion of our groups decomposed the house into "base" and "roof" sub-assemblies, completed separately and joined later. Expert users said that although team building is common on large, complex projects, closely-coupled collaboration around individual prims is rare. It is more efficient to decompose a project into sub-tasks that allow specialization and independence from synchronicity. For example, one user might create a building's skin, while another creates its furniture, a third builds roads and a fourth applies textures.

The ability for an individual to obtain multiple points of view quickly with a movable camera may make joint placing of objects redundant. On the other hand, some features of the SL UI do not support close collaboration well. For example pointing is coarse-grained, and the edit highlight on an object is seen only by the user who is editing it. When an object is being moved, only the end-point of the movement is conveyed to collaborators.

In studies of collaboration around objects, close collaboration was often "forced". [11] required one user to move a distant object while a colleague closer to it guided its placement. [12] implemented gravity so that two users were needed to lift objects while a third joined them.

It may be that VW users will only collaborate closely if real-world constraints such as gravity and strict embodiment of camera within avatar are introduced. But these constraints are not necessary in a virtual world. Rather than insisting on mimicking physical reality to enable tightly-coupled interaction, it might be more productive to embrace VWs' "unrealistic" properties.

Conclusion

It is interesting that problems that were identified more than ten years ago in early CVEs are still prevalent in a "mass market" environment like SL. In particular, difficulties with the UI (especially the lack of transparency and feedback about a collaborator's actions) still lead to a tendency to partition collaborative building into isolated sub-tasks that can be completed in parallel and assembled at the end. Our data suggests that to best support collaboration in VWs, designers should "decouple" them from physical reality to leverage their unique properties. A lot could be gained by making objects, rather than avatars, richer and more interactive.

10. REFERENCES

[1] Boellstorff, T.: Coming of age in second life: an anthropologist explores the virtually human. Princeton University Press, Princeton (2008)

[2] Carlson, J.R., George, J.F., Burgoon, J.K., Adkins, M.,White, C.H.: Deception in computer-mediated communication. Group Decision and Negotiation, 13 (1). (2004) 5-28

[3] Churchill, E.F., Snowdon, D.N.,Munro, A.J.: (eds.). *Collaborative virtual environments : digital places and spaces for interaction*. Springer, London, 2001.

[4] Dennis, A.,Valacich, J.: Rethinking media richness: Towards a theory of media synchronicity. (1999)

[5] Gibbs, M., Hew, K.,Wadley, G.: Social Translucence of the Xbox Live Voice Channel. (2004)

[6] Hindmarsh, J., Fraser, M., Heath, C.,Benford, S.: Virtually missing the point: configuring CVEs for object-focused interaction. In: Collaborative Virtual Environments: Digital places and spaces for interaction, Springer, London, (2001)

[7] Kraut, R.E., Gergle, D.,Fussell, S.R.: The Use of Visual Information in Shared Visual Spaces: Informing the Development of Virtual Co-Presence. (2002)

[8] Levinson, S.C.: Frames of reference and Molyneux's question: crosslinguistic evidence. In: Bloom, P.,Peterson, M.A. (eds.) Language and Space, MIT Press, Cambridge, (1996)

[9] Lober, A., Schwabe, G.,Grimm, S.: Audio vs. chat: The effects of group size on media choice. (2007)

[10] Moore, R., Ducheneaut, N.,Nickell, E.: Doing Virtually Nothing: Awareness and Accountability in Massively Multiplayer Online Worlds. (2006)

[11] Pinho, M.S., Bowman, D.A.,Freitas, C.M.D.S.: Cooperative Object Manipulation in Immersive Virtual Environments: Framework and Techniques. (2002)

[12] Roberts, D., Wolff, R.,Otto, O.: The Impact of Display System and Embodiment on Closely Coupled Collaboration Between Remote Users. In: Schroeder, R.,Axelsson, A.-S. (eds.) Avatars at Work and Play: Collaboration and Interaction in Shared Virtual Environments, Springer, (2006)

[13] Turkle, S.: Life on the screen: Identity in the age of the Internet. Simon & Schuster, New York (1995)

[14] Wadley, G., Gibbs, M.,Benda, P.: Speaking in character: using voice-over-IP to communicate within MMORPGs. (2007)

[15] Watson-Manheim, M.B.,Belanger, F.: Communication mode repertoires: dealing with the multiplicity of media choices. MIS Quarterly, 31 (2). (2007) 267-293

[16] Yee, N., Bailenson, J.N., Urbanek, M., Chang, F.,Merget, D.: The unbearable likeness of being digital: the persistence of nonverbal social norms in online virtual environments. Cyberpsychology and Behaviour, 10 (1). (2007) 115-121