Johannes Luderschmidt
People and Technology
Royal Holloway University of London

15. January 2008

Make Ends Meet: Collaboration Opportunities in Second Life

1 Introduction

This report presents an overview of Collaborative Virtual Environments (CVE) as a field of Computer Mediated Collaboration. CVE research activities will be introduced and their findings will be described. CVEs have been intensively researched since the mid 1990s. Second Life (SL) seems to be the first commercial Virtual Environment that has been adopted by a large community. As of September 2007 according to the company behind SL Linden Labs more than 9.5m users have registered for SL (Linden Labs, 2007 [1]). SL is a Virtual Environment respectively Virtual World where users interact with other users, objects or agents by means of a client called viewer on their computer that communicates with a central server grid. Second Life is a *consumer* VE that offers multiple possibilities for collaboration. This report applies findings of CVE research activities to Second Life and introduces some ideas how CMC activities like virtual meetings can be realised with Second Life. SL's current monopoly on the field of Virtual Environments promises to be a salient platform to research and implement CMC activities on. In terms of successful CVE implementations it seems to be more likely that lacks in SL can be overcome to realise collaboration activities than that a VE environment solely for collaboration purposes will establish in the near future.

2 Collaborative Virtual Environments

Collaborative Virtual Environments address CMC aspects in Virtual Environments. A VE is a special field of Virtual Reality. It is a graphically rich, three-dimensional Virtual World where (geographically dispersed) users usually in form of an avatar interact with other users, objects or agents. In contrast to Virtual Reality systems it is not necessary for the presentation of VEs to "block out the world" by using displays like Head Mounted Displays or CAVEs to "present virtual-objects in life size" and therefore "create the illusion of immersion" (Brooks, 1999, [2]). Thus VEs can be used on standard screens. CVEs com-

bine sound, text and gestures in Computer Mediated Communication Processes. They are groupware that utilize CMC capabilities like chatting and combines them with social and spatial aspects. E.g. when not collocated in the same physical space psychological dimensions of a person's virtual embodiment and social presence can be expressed in CVEs (Verhulsdonck, 2007, [3]).

2.1 Basic Findings

Important research is based on MASSIVE systems of the University of Nottingham introduced by Greenhalgh and Benford in [5]. MASSIVE was intended as a prototype Virtual Reality teleconferencing system that has been used to research CVEs. Bowers et al used it in 1996 for their research ([4]) and came to following foundational results:

- Participants interaction in real world impacts upon interaction in Virtual Worlds.
- An avatar needs to have a shape that has certain human like attributes for instance a user needs to know the body axis's just like Tversky et al described in their paper ([7]): front and back, a vertical axis (between feet and head) and a horizontal axis (left hand to right hand). Bowers et al used cross-shaped bodies called Blockies consisting of a vertical box onto which a horizontal box was stacked. Additionally a square was used on the horizontal box to indicate eyes and thus making clear that this side is the front. A detailed treatment of "User embodiment in collaborative virtual environments" is available from Benford et al ([8]).
- It is not necessary that a Virtual World is foto realistic.
- Users try to align their avatars in conversations in a way that the front faces the partner in conversations in order to install a face-to-face situation with the partner.
- Comments like "hi i'm here behind kai" show that users are spatially aware of other users.
- An embodiment must be trustable. For instance the avatar of a user who suffered technical failure must not become a zombie.
- Social interaction should be the main concern when designing a Virtual World. Contrary, "thrilling aesthetics" are not crucial. E.g. a meeting table should not be missing in a virtual meeting room. It can help users to establish a joint orientation in order to convey communicational information.

- Subtleties of social interaction calls for supplementary channels of information like gestures or facial expressions. In the case of MASSIVE interaction a similar user behaviour like in videoconferencing was recognizable: If something was not visible to a user it was compensated by using the audio channel to replace missing functionality.

- A body must provide means to look around without changing the body's orientation. There is an important difference in body language between looking around and turning away. The implementation of a VE should reflect that issue.

In 2000 Hindmarsh et al researched how object-focused interaction can be realised in CVEs [6]. Therefore they used the MASSIVE 2 systems that ran on a Silicon Graphis Workstation. Users should carry out certain object related tasks e.g. one user should move a lamp in a virtual room. Where the lamp should be positioned was told by another user (observer) in that room. Avatars could use deictic gestures (e.g. pointing gestures) as well as grasp objects and users could use an audio connection. They figured out that due to a 55-degree field of view of the 3D scenes a fragmented workspace emerged. It was difficult for the observer to see the pointing gesture of an avatar as well as the object that was pointed at. Due to that the audio connection was used to explicitly describe which object was meant.

Because of delays it was not clear where another avatar was precisely positioned. For instance an avatar B could stand next to user A's avatar on screen but the user that controlled avatar B already had turned around its avatar on its own screen. Thus voice explanations like "I mean the TV you're looking at" did confuse the other user.

Handling and movement of avatars in CVEs may be awkward without training of users because users need to get used to navigation metaphors. Also it is important that certain actions can be accomplished parallel e.g. pointing and looking around at the same time to find another user. These actions were impossible in MASSIVE 2 (Hindmarsh et al, 2000, [6]).

2.2 Communication Channels

In CVEs users are allowed to communicate with others via various communication channels e.g. in SL:

- Textual: Chat, Instant Messages

- Auditory: VoIP

- Visual: Facial expressions/ gestures
- Symbolic: Manipulation of objects
- Spatial: Avatar Movement
- Deictic: Pointing

VEs can potentially overcome the loss of physical presence with the support of visual communication like body language, hand gestures or facial expressions that yield social cues. CMC applications such as instant messaging or audio conferences do not bear those social cues. Contrary, videoconferencing supports facial expression and body gestures. However, it does not offer to share the same space and does not allow an easy turn taking of speakers or deictic gestures such as to point at other objects. A phenomenon of media with a lack of social cues is that people feel less responsible for their behaviour because they are not in the same place as their communication partner.

It is equally important *what* somebody says as *how* somebody says it. The *how* is based on a medium's communicative channels (Verhulsdonck, 2007, [3]). Gestures may support CVE users in gathering common ground because they allow non-verbal exchange that humans are used to in face-to-face communication. For VEs it would be useful to design strong gestures (deictic gestures like pointing) and less-important contextual gestures for special purposes (e.g. iconic, metaphoric, beat/cohesive gestures). Iconic gestures are those indicating a special purpose e.g. an arm stretched and a thumb stuck out indicates that someone is looking for a hitchhiking possibility. Metaphoric gestures are trying to give visual support to abstract aspects e.g. closing hands to indicate that a part of a speech is coming to an end. Beat/cohesive gestures are quite similar but they have an emphasizing nature. For instance a politician that waves hands with open palms up and down wants to emphasize a point. However, iconic, metaphoric and beat/cohesive gestures are more complicated to design. E.g. for turn-taking an avatar could raise its hand combined with a playback of a "sorry" sound. This gesture could involve different kinds of intensity (friendly and muted, less friendly and muted, interruptive etc.) to give more information on the listener's mood who wants to take turn. Contrary, deictic gestures should be easier to design and they are presumably more important to establish common ground and thus more beneficial in conventional CVE communication (Verhulsdonck, 2007, [3]).

2.2.1 Compare Videoconferencing with Audio and CVE

Clayes and Anderson accomplished a study in which groups of three people had to collaborate in four different tasks. They either saw video images of the two others or images

with static avatars over which the names of the two other participants [9] were hovering. The static avatar images contained no personal information about the other person. Just one of two predefined images dependent of the gender of the collaboration partner was used. Thus the Virtual Environment was really basic. Face tracking was employed to trace where the participants where looking at during the tasks. Afterwards the participants had to fill out a short questionnaire. Results of the study:

- The context of communication is important: It is more important to see one's conversational partner personally in social tasks compared to problem solving tasks.
- Video images provide visual feedback and thus facilitate turn taking i.e. video images may be beneficial for remote users in terms of reducing the number of interruptions. Users tend to look more often at videos than to avatar icons. Nonetheless, participants used to look at a certain avatar when they addressed the other participant represented by this avatar.
- Video images may be effective in terms of reducing the number of interruptions. In the questionnaires they were not perceived to be more sociable, warm or sensitive compared to avatar icons.
- Avatar representations enable participants to achieve the same level of task performance (i.e. problem solving tasks were completed equally fast and equally precise). Furthermore, avatar representations were rated equally to video representations.

2.3 Media Channels in VEs and Psychology

Media Naturalness Theory (MNT) tries to explain why humans prefer certain natural media channels for communication. Humans favour face-to-face synchronous communication that supplies immediate response through aural and visual information (eye contact, body language, gestures) and stimulates physiological senses through proximity to other humans' presence in the same place. Nevertheless, according to MNT it is possible to adapt to media like email and instant messaging also, but this comes at a greater expense. It is more demanding for our communicative abilities as we must iteratively learn to overcome and compensate for a medium's limitations. A medium's "naturalness" can be measured with a communicative expense to which we can adapt with added perceptual costs. With time and effort a human can accustom to every medium that corresponds to natural practices. Contrary, "richness" of a medium does not necessarily mean that it

will be adopted by users. According to MNT, face-to-face communication should be approximated by CVEs' communicative channels to be more natural (Verhulsdonck, 2007, [3]).

3 SL as a platform for CVE

Second Life is a Virtual World into which a user can log in with a certain client application called viewer. This viewer is available for Microsoft Windows, Mac OS X and Linux operating systems and can be downloaded for free from the Internet. Before a user can log in to SL for the first time it must be registered for free on secondlife.com. In the Virtual World of Second Life a user's avatar can own and use clothing, objects, gestures, sounds etc., which are available via an inventory in the viewer. SL has its own currency called Linden Dollar (abbr. L$) that can be used for payment within the boundaries of SL.

3.1 Collaboration possibilities

Community or identity building issues are interesting activities in Second Life but these may not be relevant in terms of CVE research. Second Life offers technological possibilities for social interaction that should be taken into consideration for CVE purposes.

Second Life is extensible by means of scripts, objects, gestures, sounds etc. and therefore offers a development platform for CVE applications. SL provides simple means for interaction e.g. right-clicking on a chair offers a choice to "Sit here", right clicking on an avatar offers "Send Instant Message" etc. Second Life consists of 256x256 m regions. A user logs in to a region and can afterwards change regions by teleporting from one region to another. Inside of a region there are different means of communicating with each other. A common method is to use the chat engine. A user just enters some text in the text field of its viewer and presses enter. This text will appear in the viewers of all avatars that remain within a certain surrounding. Additionally instant messages can be sent from one avatar to another without being visible to other users.

Second Life offers inherent possibilities that videoconferencing does not offer so far. Voice over IP was introduced in 2007 and offers 3D sound. According to secondlife.com 'Second Life voice uses 3-D "proximity-based" spatial awareness, taking distance and direction into account, for a more realistic in-world voice experience. When speaking, gestures can animate avatars as their voices become louder' (2007, [13]). I.e. if multiple persons talk at one time the voices of users will have different loudnesses depending on the distance to one's avatar. This enables for instance side conversations which would

not be possible between two partners in a videoconference. Additionally pointing is possible with deictic gestures: Avatars can raise their arms and point at objects to guide the gaze of another user. Mouse-moves-head takes care that the conversation partner a user is looking at sees that this user is gazing at him, which inherently offers capabilities that Werkhoven et al called "isotropic view of participants" (2001, [14]). In videoconferencing offering an isotropic view means that every user has to be displayed in a single window. According to Werkhoven et al this method yields better persuasive force and dominance. However, there are videoconferencing systems that offer the feature of isotropic views (e.g. Apple's *iChat* [15]) but the amount of remote participants in a videoconference is restricted. E.g. in iChat up to four persons can use isotropic videoconferencing. Another feature in SL is the built in chat-engine, which allows global chatting or sending of Instant Messages privately from one user to another. This engine can be used for additional information exchange between participants e.g. while someone is giving a presentation in SL two (or more) users can exchange for instance about presentation contents without disturbing the presenting user.

3.2 Critical Mass

Why had there been no publicly successful implementation of a CVE in the intensive research period around 2000? There seems to be at least two important aspects to this problem. One aspect surely are technological difficulties. MASSIVE 2 client software needed an expensive Silicon Graphics workstation in the year 2000. On the other hand broadband Internet connections were quite rare.

SL has overcome those obstacles and reached a critical mass of users as the amount of more than 9.5 million avatars that has been registered [1] gives proof. There are different reasons for that. Firstly, Second Life is not expensive as it can be downloaded and used for free ([10]). The client application is even Open Source ([11]). Everyone can instantly create its own avatar on *secondlife.com* and log into SL. Secondly, SL can be used on (cheap) consumer hardware.

It also seems like Second Life developers considered CVE research results from section 2.1. Interaction possibilities are based on real world interaction. Avatars are (mostly) human like creatures that can use pre-set gestures involving facial expressions, hand gestures, body language combined with sounds and text chat. These channels approximate natural communication and according to Media Natural Theory (see section 2.3) this should increase the possibility that they will be adopted by users. Second Life is fairly stable and if a viewer crashes it will not leave a zombie avatar. Object building possibil-

ities were sophistically implemented to prevent a fragmented workspace. E.g. an avatar can edit an object and move its head parallelly to see what's happening around it (see section 2.1). Every avatar in the surrounding can see which object is being edited because one arm of the editing avatar is stretched out to the edited object and white dots form a line between the avatar's hand and the object itself. If the user moves its mouse to look around in the 3D scene the corresponding avatar will move its head and look in the direction where the mouse pointer is (mouse moves head). Hence if a user moves the mouse pointer to the face of another avatar its own avatar will automatically look at the other one. This makes instantly clear who actually is socially engaged with whom. By doing so SL enables that an avatar makes a difference between turning the head to someone and turning the back on someone.

3.3 Virtual meetings and object focused interactions

SL offers possibilities to build shared artefacts e.g. canvases for presentations. These are implemented in SL as embedded video views displayed on dynamic texture maps. Videos can be displayed on objects e.g. canvases via streaming media. This could be used to give a Powerpoint presentation. Alternatively, dynamic texture maps can be configured as embedded browser views to display web pages visible to every participant. Other shared artefacts could be build with the built-in scripting language *Linden Scripting Language* (LSL), which allows users and developers to add behaviours, special effects or Internet communication capabilities to objects. E.g. artificial intelligence can be added to an object in order that it behaves like a pet that follows a user around or to dynamically load a 3D model from the Internet. LSL can also be used to create head-up displays (HUDs) that display additional information. HUDs in SL are used to show information directly in the window of the viewer software (like GUI elements) instead of projecting it into the 3D scene. As the source code of the viewer is published under an Open Source license developers can download and modify the source code of the viewer and build advanced features for collaboration into the existing software. This could for instance be employed to build a CAD viewer (e.g. explained by Rosenman et al [16]) that could read and display files from applications like AutoCAD. Architects could meet in Second Life to discuss progress of work directly by showing corresponding CAD models. A HUD could be used to display information about selected parts of the CAD model. However, in this scenario every user would need the modified version of the viewer software that offered the additional CAD functionality.

4 SL CVE Research Opportunities

In section 3.3 some ideas about virtual meetings and object focused interaction had been introduced. It would be interesting to create and supervise different collaboration scenarios in SL to gain insight to collaboration issues. How could and would a typical meeting take place in Second Life: Would there be one user giving a Powerpoint presentation and the rest is quietly listening/watching? Or would participants e.g. make use of the Instant Messaging function and send messages to other users while the presentation? When would a meeting in SL make sense? A hypothesis could be that a meeting in SL would only makes sense if there were many participants in different locations. If there were participants dispersed over two locations a videoconference could be preferred by participants. Many users seem to dislike or to find it silly to use a Virtual Reality application because e.g. they are of the opinion that the goal of Virtual Reality is to render real life useless. This fear could be an obstacle to get immersed into a 3D Virtual World like SL and to successfully participate in a virtual meeting. This all could be tested and checked with questionnaires after the tests.

How could object-focused interaction be implemented? Second Life already offers manipulation gestures like those demanded by Hindmarsh et al ([6]). But would they be employed by participants? How could access to gestures and facial expressions be provided 'intuitively' to the users? Which gestures/facial expressions would be demanded as missing. How would turn-taking take place? Via a default gesture like a raised hand, via voice or with the global chat engine (e.g. an avatar writes "Sorry, I've got a question.")? How would the voice implementation in SL compare with voice in videoconferencing especially if there are many participants talking at the same time? How distracting would be the navigation in a 3D environment? Would the additional complexity of controlling an avatar in SL seriously damage collaboration? How would experienced users perform in relation to users new to Second Life/3D environments?

Psychologically it would be interesting to test the rate of information exchange and persuasive force. This could e.g. be figured out with experiments like *The lost at the Moon Task* and *The Case of the Fallen Businessman* (see [14] for further details).

5 Summary And Conclusion

5.1 Summary

CVEs have long been an issue for collaboration research. Basic research has been performed on MASSIVE VEs from the University of Nottingham. It has been figured out that users of VEs have a spatial awareness of their partners and try to get face-engaged when communicating with them via avatars. Social interaction must be subtly implemented in a VE e.g. it must be possible to make a difference between turning an avatar's head or body. If something is difficult to explain or cannot be shown with methods of the VE users tend to explain what they do via audio. Additionally users need to exercise how to navigate in a CVE. Compared to collaboration methods like videoconferencing VEs offer additional communication channels e.g. deictic gestures, spatial communication by moving an avatar or symbolic communication by manipulating objects. Communication channels in CVEs are quite natural because they approximate communication channels that humans also employ in face-to-face communication. According to Media Naturalness Theory natural communication channels will be faster adopted than useful but unnatural channels.

Second Life is an example for a 3D Virtual Environment that offers a vibrant community and instead of research projects like MASSIVE SL is a platform used by a large community. Many collaboration possibilities are already implemented or can be added (see section 3.3). SL offers collaboration research possibilities e.g. those introduced in section 4.

5.2 Conclusion

With the success of SL there are new opportunities to realise a CVE that can be used by collaboration partners around the globe. The idea is to use SL as a technical layer for CVE. The given, free functionality and expandability can be employed to implement special CMC functionalities on top of it. Meetings in SL could help to bring a lot of users from geographically dispersed locations together on one (virtual) table. Every user in SL has an isotropic view on other partners. In contrast to videoconferencing users can see where another user's gaze is directed at. In videoconferencing systems persons that talk parallelly are a problem because voices of participants usually have the same loudness. 3D sound in Second Life seem to be very promising. Users have spatial hearing impressions of other users talking around them. This could e.g. enable side conversations in meetings. Users subconsciously recognise spatial communication and the embodiment of meeting

partners influences their behaviour. However, when users start to use SL they could be distracted by navigation difficulties.

A virtual meeting in Second Life will never make a face-to-face meeting obsolete. But it could be a useful alternative to videoconferencing with many participants in different locations.

References

[1] Linden Labs, 2007, *stats_200710*, Available from: <http://static.secondlife.com/economy/stats_200710.xls> [Accessed 3. January, 2008]

[2] Brooks Jr., F. P. 1999. *What's Real About Virtual Reality?*, In IEEE Computer Graphics And Applications, 19(6), 16

[3] Verhulsdonck, G. 2007. *Issues of Designing Gestures into Online Interactions: Implications for Communicating in Virtual Environments*, In SIGDOC'07, October 22-24, 2007, El Paso, Texas, USA

[4] Bowers, J., Pycock, J. and O'Brien, J. 1996. *Talk and embodiment in collaborative virtual environments*, In CHI '96: Proceedings of the SIGCHI conference on Human factors in computing systems, 1996, Pages 58-65

[5] Greenhalgh, C. and 1995. *MASSIVE: a collaborative virtual environment for teleconferencing*, In ACM Trans. Comput.-Hum. Interact., 1995, Pages 239–261

[6] Hindmarsh, J., Fraser, M., Heath, C., Benford, S. and Greenhalgh, C. 2000. *Object-focused interaction in collaborative virtual environments*, In ACM Trans. Comput.-Hum. Interact., 2000, Pages 477–509

[7] Tversky, B., Bauer Morrison, J., Franklin, Bryant, D.J. 1999. *Three Spaces of Spatial Cognition*, In The Professional Geographer, November 1999, Volume 51, Pages 516-524(9)

[8] Benford, S., Bowers, J., Fahlén, L.E., Greenhalgh, C. and Snowdon, D. 1995. *User embodiment in collaborative virtual environments*, In CHI '95: Proceedings of the SIGCHI conference on Human factors in computing systems, 1995, Pages 242–249

[9] Clayes, E.L. and Anderson, A.H. 2006. *Real faces and robot faces: The effects of representation on computer-mediated communication*, In International Journal of Human-Computer Studies, 2007, Volume 65, Issue 6, Pages 480–496

[10] secondlife.com, 2007, *Second Life Community: Downloads*, Available from: <http://secondlife.com/community/downloads.php> [Accessed 10. January, 2007]

[11] Second Life Wiki, 2007, *Source downloads - Second Life Wiki*, Available from: <http://wiki.secondlife.com/wiki/Source_downloads> [Accessed 30. December, 2007]

[12] Francik, E., Ehrlich Rudman, S., Cooper, D. and Levine, S. 1991. *Putting innovation to work: adoption strategies for multimedia communication systems*, In Commun. ACM, Volume 34, Number 12, 1991, Volume 51, Pages 52–63

[13] Secondlife.com, 2007, *The Technology behind the Second Life Platform*, Available from: <http://s3.amazonaws.com/download.grid.secondlife.com/Fact_Sheet_Technology.pdf> [Accessed 12. December, 2007]

[14] Werkhoven, P.J., Schraagen, J.L., Punte, P.A.J. 2001. *Seeing is believing: communication performance under isotropic teleconferencing conditions*, In Displays, 2001, Issue 22, Pages 137–149

[15] Apple, 2008, *Apple - Mac OS X Leopard - Features - iChat*, Available from: <apple.com/macosx/features/ichat/> [Accessed 14. January, 2008]

[16] Rosenman, M.A., Smith, G., Maher, M.L., Ding, L., Marchant, D. 2007. *Multidisciplinary collaborative design in virtual environments*, In Automation In Construction, 2001, Issue 16, Pages 37–44