Voice in virtual worlds

Gregory Robert Wadley

Submitted in total fulfilment of the requirements of the degree of Doctor of Philosophy.

December, 2011

Department of Information Systems,
Faculty of Science,
The University of Melbourne
Abstract

Virtual worlds are simulated online spaces through which large numbers of people connect in order to work, play and socialize. Examples include massively multiplayer online games like World of Warcraft and open-ended worlds like Second Life. Virtual worlds are differentiated from other systems by their simulation of a persistent three-dimensional landscape, in which users are usually represented as avatars. Virtual worlds host large numbers of users and support a variety of recreational and instrumental uses.

A critical aspect of any collaborative system is communication. In virtual worlds this is especially complex. Most users encounter people who are unknown to them offline. Most maintain a presence in the world over many months or even years, yet may prefer to be pseudonymous or engage in identity-play. Users must simultaneously manage both physical and virtual contexts. Synchronous as well as asynchronous communication is required.

Virtual worlds initially offered only text as a medium for user-to-user communication. More recently, vendors have introduced facilities for communicating by voice. This has made the experience of virtual worlds more convivial for some, and has enabled forms of collaboration that were previously only possible in small experimental systems. But the introduction of voice provoked controversy, with some protesting that it projects too clearly the personal characteristics of speakers, damaging pseudonymity. Some are more sensitive about speaking with strangers than they are typing, and may become less communicative when adopting voice, or more easily dominated by extroverted collaborators. Voice channels are more prone to abuse, and the abuse can be more impactful. Users encounter sound quality problems, and are often uncertain whether they are being heard. Voice is less suited to asynchronous communication, and is more prone to congestion.

It appears that voice works well when conditions suit, but can lead to failed implementations when deployed inappropriately. Yet little research has been conducted to help us understand to which situations voice is suited, whom it benefits and whom it does not, and how it can best be configured to support different activities.

I conducted four studies designed to fill this gap. Two examined the influence of voice on user experience, in the two major types of virtual world. The others examined the interaction between voice and spatiality, the defining feature of virtual worlds, at macro and micro scales. I studied use in naturalistic contexts, collecting data via interviews and diaries and
triangulating these with observation, online ethnography, conversation analysis and quantitative measures.

I found that voice transforms the user experience of virtual worlds. It makes some forms of collaboration more efficient. However it interferes with identity-play and the ability of users to manage multiple tasks and conversations. When voice is propagated spatially, it increases immersion, reduces channel clutter, and affords new strategies for team coordination. However verbal references to places and objects often fail.

I discuss these results in the light of post-media-richness theories of communication, arguing that preferences for one modality or another reflect broader issues of managing social presence in virtual and physical contexts.
Declaration

This is to certify that:

(i) the thesis comprises only my original work towards the PhD except where indicated in the Preface,

(ii) due acknowledgement has been made in the text to all other material used,

(iii) the thesis is fewer than 100,000 words in length, exclusive of tables, maps, bibliographies and appendices.

Signed,

______________________________

Greg Wadley

Dated: ________________________

Preface

Studies 1 and 2 of this thesis were conducted in collaboration with Martin Gibbs and Peter Benda at the University of Melbourne. Martin and I collaborated in designing these studies, and all of us collaborated in collecting the data; however the data analysis and write-up presented in this thesis was conducted by me.

Study 3 was supervised by Nic Ducheneaut and conducted by me.

Study 4 was supervised by Martin Gibbs and conducted by me.

All four studies have been presented at conferences. The publications are listed in Appendix C. These were multi-authored; however the text in this thesis was written by me.
Acknowledgements

A thesis cannot come into being without the help and influence of many people, and I would like to thank them now. I make the usual author’s disclaimer that the fault for any inadequacies in the work lies with me alone.

My supervisors Martin Gibbs and Steve Howard of the University of Melbourne steered me deftly through a long and sometimes arduous process.

The Interaction Design Group at the University of Melbourne has offered unceasing challenge and stimulation while I worked on this thesis, and I thank all past and current members.

I have worked with the staff of the Department of Information Systems at Melbourne for a decade and I thank them for their support and collegiality.

I learned by working with experienced researchers: thanks Pete Benda, Nic Ducheneaut and Martin Gibbs.

I spent four months researching at the Palo Alto Research Center in California, funded by a University of Melbourne PORES scholarship. This was a transformative experience and I thank everyone who made it possible, especially staff at PARC, the Department of Information Systems, and the University of Melbourne’s School of Graduate Research. Most of all I thank my supervisor at PARC, Nic Ducheneaut, and my fellow researchers there Don Wen and Mike Robinson.

Many people discussed my work with me and offered useful advice and insights, including Richard Bartle, Tom Boellstorff, Marcus Carter, Paul Dourish, Connor Graham, Mitchell Harrop, Jeremy Hunsinger, Reeva Lederman, Tuck Leong, Bernd Ploderer, Ralph Schroeder, Dmitri Williams, and fellow PhD candidates in the Interaction Design Group.

I thank Paul Boustead for making the SpatialVoice system available for study 2.

I would like to thank the people who participated in my studies. Study participants receive little compensation for their efforts, and research cannot happen without them.

Finally thanks to my friends and family for their support.
# Table of Contents

Voice in virtual worlds................................................................................................................... i

1. Introduction .............................................................................................................................. 1

1.1 Background .......................................................................................................................... 1

1.2 Classification and terminology ........................................................................................ 7

1.3 Theme: Communication in virtual worlds ................................................................. 13

1.4 Problem ............................................................................................................................. 15

1.5 Prior work .......................................................................................................................... 16

1.6 Gap ................................................................................................................................... 17

1.7 Research Question............................................................................................................. 17

1.8 Approach............................................................................................................................ 17

1.9 Outcomes and contributions............................................................................................ 20

1.10 Conclusion ....................................................................................................................... 21

2. Review of related research ...................................................................................................... 23

2.1 Introduction ....................................................................................................................... 23

2.2 Virtual worlds ................................................................................................................... 23

2.3 Mediated communication................................................................................................. 44

2.4 Voice in virtual space......................................................................................................... 59

2.5 User experience.................................................................................................................. 62

2.6 Gap .................................................................................................................................... 65

2.7 Research question............................................................................................................. 65

2.8 Conclusion ......................................................................................................................... 67

3. Design of my research program .............................................................................................. 69

3.1 Introduction......................................................................................................................... 69

3.2 Phenomena and challenges .............................................................................................. 69
3.3 Overall approach ............................................................................................................ 75
3.4 Selection of cases ........................................................................................................... 81
3.5 Methods used in individual cases ................................................................................... 83
3.6 Presentation of results .................................................................................................... 87
3.7 Conclusion ..................................................................................................................... 88

Study 1: Voice in massively multiplayer online games .............................................................. 89
4.1 Background .................................................................................................................... 89
4.2 Method ........................................................................................................................... 93
4.3 Findings ........................................................................................................................ 98
4.4 Discussion .................................................................................................................... 135
4.5 Conclusion and next steps ............................................................................................ 143

Study 2: Spatially-propagated voice ......................................................................................... 145
5.1 Background .................................................................................................................. 145
5.2 Method ......................................................................................................................... 147
5.3 Findings ........................................................................................................................ 150
5.4 Discussion .................................................................................................................... 162
5.5 Conclusion and next steps ............................................................................................ 167

Study 3: Communication during collaboration around objects ................................................. 169
6.1 Background .................................................................................................................. 169
6.2 Method ......................................................................................................................... 177
6.3 Findings ........................................................................................................................ 184
6.4 Discussion .................................................................................................................... 193
6.5 Conclusion and next steps ............................................................................................ 198

Study 4: Voice in a social virtual world .................................................................................... 199
7.1 Background .................................................................................................................. 199
7.2 Method ......................................................................................................................... 201
7.3 Findings ........................................................................................................................ 205
List of tables and figures

Tables

Table 1: Metaphors for propagating voice and text messages .................................................... 14
Table 2: Overview of my four case studies ................................................................................. 20
Table 3: Data collection methods used in each study ................................................................. 87

Figures

Figure 1: Using text-chat in the Second Life virtual world........................................................... 2
Figure 2: Timeline of virtual environment technologies............................................................... 4
Figure 3: Relationship between CVEs and other technologies ..................................................... 9
Figure 4: Taxonomy of virtual environments.............................................................................. 12
Figure 5: Cases arrayed along dimensions of diversity............................................................... 82
Figure 6: My cases situated within taxonomy of virtual environments ....................................... 82
Figure 7: Design of groups for study 1 ....................................................................................... 95
Figure 8: Data collection timeline for study 1............................................................................. 95
Figure 9: Text vs voice in MMOGs .......................................................................................... 138
Figure 10: Data collection timeline for study 2......................................................................... 150
Figure 11: In-avatar and in-camera views................................................................................. 173
Figure 12: Creating prims in Second Life ................................................................................... 174
Figure 13: 'House' task screenshot provided to participants ..................................................... 178
Figure 14: 'Garden' screenshot provided to participants ............................................................ 180
Figure 15: Data collection timeline for individual trials in study 3 ............................................ 181
Figure 16: Split-screen video of a trial in study 3 ....................................................................... 183
Figure 17: Time spent with camera in in-avatar mode............................................................... 189
Figure 18: Frames of reference used to refer to objects in speech ............................................ 191
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 19</td>
<td>Data collection timeline for study 4</td>
<td>203</td>
</tr>
<tr>
<td>Figure 20</td>
<td>Encountering voice users in SL (my avatar in foreground)</td>
<td>204</td>
</tr>
<tr>
<td>Figure 21</td>
<td>PLS 1 for study 1</td>
<td>275</td>
</tr>
<tr>
<td>Figure 22</td>
<td>PLS 2 for study 1</td>
<td>276</td>
</tr>
<tr>
<td>Figure 23</td>
<td>Interview questions for study 1</td>
<td>277</td>
</tr>
<tr>
<td>Figure 24</td>
<td>Diary instructions for study 1</td>
<td>278</td>
</tr>
<tr>
<td>Figure 25</td>
<td>First pass at code hierarchy for study 1</td>
<td>282</td>
</tr>
<tr>
<td>Figure 26</td>
<td>Code hierarchy for study 1</td>
<td>283</td>
</tr>
<tr>
<td>Figure 27</td>
<td>Participant instructions for study 2</td>
<td>284</td>
</tr>
<tr>
<td>Figure 28</td>
<td>Coding hierarchy from study 2</td>
<td>286</td>
</tr>
<tr>
<td>Figure 29</td>
<td>IRB application for study 3</td>
<td>288</td>
</tr>
<tr>
<td>Figure 30</td>
<td>Participant recruitment advertisement for study 3</td>
<td>289</td>
</tr>
<tr>
<td>Figure 31</td>
<td>Study 3 lab notes 1</td>
<td>290</td>
</tr>
<tr>
<td>Figure 32</td>
<td>Study 3 lab notes 2</td>
<td>291</td>
</tr>
<tr>
<td>Figure 33</td>
<td>Study 3 lab notes 3</td>
<td>292</td>
</tr>
<tr>
<td>Figure 34</td>
<td>Discussion about building in Second Life</td>
<td>293</td>
</tr>
<tr>
<td>Figure 35</td>
<td>Sample video analysis: one pair</td>
<td>294</td>
</tr>
<tr>
<td>Figure 36</td>
<td>PLS for study 4</td>
<td>295</td>
</tr>
<tr>
<td>Figure 37</td>
<td>Interview questions for study 4</td>
<td>296</td>
</tr>
<tr>
<td>Figure 38</td>
<td>Excerpt from my SL diary</td>
<td>297</td>
</tr>
<tr>
<td>Figure 39</td>
<td>Excerpt of a group discussion from study 4</td>
<td>298</td>
</tr>
<tr>
<td>Figure 40</td>
<td>Excerpt of an interview from study 4</td>
<td>299</td>
</tr>
<tr>
<td>Figure 41</td>
<td>Excerpt from Second Life user forum</td>
<td>300</td>
</tr>
<tr>
<td>Figure 42</td>
<td>Data coding hierarchy from study 4</td>
<td>301</td>
</tr>
</tbody>
</table>
Chapter 1

1. Introduction

1.1 Background

This thesis is concerned with how people communicate within virtual worlds: large, Internet-based systems through which geographically-dispersed people connect for recreation, social interaction and collaborative work. Well-known virtual worlds (the term is usually abbreviated as VWs1) include Second Life and World of Warcraft. VWs are a subset of the genre of computer systems known as collaborative virtual environments (CVEs). CVEs simulate three-dimensional spaces and visualize them using real-time computer graphics. Virtual spaces are usually designed to look something like the real world of our everyday experience, though they are typically fictional.

By definition, a collaborative virtual environment is a multi-user system (Churchill et al., 2001). Users are typically represented to each other in the form of simulated bodies called ‘avatars’. Most virtual environments of interest today are multi-user, therefore in this thesis I sometimes omit the word ‘collaborative’ for the sake of brevity and simply say ‘virtual environment’.

The properties that make VWs ‘worlds’ – which are usually accepted as differentiating virtual worlds from the larger category of collaborative virtual environments - are:

1. VWs are persistent (the simulation continues to run on the host computer, regardless which particular individuals happen to be logged in).

2. VWs are used by large numbers of people (World of Warcraft, a popular system, currently has about 11 million users).

3. The space of a VW is large (for example a broad expanse of virtual land, as opposed to, say, a single room).

1 See appendix A for typographical conventions and appendix B for a glossary of terms and acronyms.
Like most Internet-based social platforms, virtual worlds typically allow their users to communicate with each other by typing text messages. More recently however, many have added voice communication facilities, so that now both written and spoken forms of verbal communication are available. In a voice conversation, sound recorded by the sending user’s microphone is transmitted over the Internet and replayed through the receiving users’ speakers or headphones. The problem of how to design successful voice systems for virtual worlds is a theme of this thesis.

Virtual world users can communicate non-verbally as well. The use of avatars affords embodied forms of communication such as proxemics (how people position their bodies in relation to each other) and a simplified language of posture and gesture. Some virtual worlds allow users to construct objects, which are visible to others and can communicate meaning.

Figure 1 shows a screenshot taken in the popular virtual world Second Life. My avatar is in the foreground, and the avatars of other users are scattered around the scene. Avatars are labelled with their user’s pseudonym, and optionally the name of a group to which the user belongs. In this scene some users have formed their avatars into conversation groups, as we would our bodies in offline conversation. Text messages entered by users whose avatars are near mine are displayed in the chat window on the bottom left. The ‘mini map’ on the top right is a top down view of the local area, with users represented as green dots. (This screenshot was taken in 2007: the SL ‘Viewer’ or client software has changed somewhat since then.)
Researchers have been examining social interactions in collaborative virtual environments for over twenty years (e.g. work surveyed in Churchill et al., 2001; Schroeder, 2002; and Schroeder and Axelsson, 2006). A significant concept in earlier work was presence, defined as the sensation of being somewhere other than one's physical surroundings; such as in a distant or fictitious place. The experimental systems examined in early research often utilized input-output (I/O) hardware beyond what is typically found on personal computers. For example Virtual Reality (VR) systems use head-mounted displays, data gloves and other specialized devices in order to maximize their users’ sense of being immersed in a virtual environment. VR did not achieve widespread commercial success, and vendor interest (and some but certainly not all researcher interest) shifted towards systems based on consumer hardware. Since the late 1990s, ‘desktop’ virtual worlds have allowed users to connect via PCs and the Internet to simulated spaces running on large servers. Some of these have been commercially successful and have inspired significant research interest.

Today’s VWs are often classified into two categories: those that are games and those that are not. The distinction is based not on the underlying technology (which is common to both) but on the kinds of activities that their users undertake. Use of a game world is constrained by rules and the necessity that users compete with each other, or collaborate to defeat simulated enemies. Activity in non-game worlds is less restricted, though there are still rules, such as the property laws in Second Life which support a market for user-created virtual goods. The taxonomy of Schultze and Rennecker (2007) classifies VWs along two dimensions: fantasy-realism and progression-emergence. These authors’ progression-emergence dimension describes the degree to which user behaviour is shaped by game rules, and thus corresponds to the distinction between games and non-games.

Game-based virtual worlds are usually called Massively Multiplayer Online Role-Playing Games, usually abbreviated as MMORPG, MMOG or just MMO. For brevity, and because the existence and nature of role-play in these worlds is contentious (e.g. MacCallum-Stewart and Parsler, 2008), I call them Massively Multiplayer Online Games (MMOGs) in this thesis.

Non-game worlds are usually called either “open ended” (to emphasize that activity is not constrained by game rules), or “social”. The term “social world” is not a reference to Anselm Strauss’ concept of the same name, but simply emphasizes that a principle activity in this genre of VWs is socializing with other users. The term “social” is problematic because it distinguishes these platforms neither from game worlds nor non-spatial systems such as social network websites, all of which also feature social activity. However I have chosen to use this term because it is briefer than the alternative “open-ended”, and is in more common usage.
Lehdonvirta (2010) has pointed out that the ‘MMOG vs Social’ categorization is problematic because it lacks a strict definition. However its popularity suggests that it captures a distinction which is important to users.

MMOGs offer large, persistent, usually fictional settings in which groups of users collaborate and compete in a long-term game. MMOGs inherit much of their culture from the earlier, non-graphical ‘multi-user dungeons’ or MUDs. Currently the most popular MMOG in the USA, and one of the most popular worldwide, is *World of Warcraft*, which has about 11 million users. Other prominent MMOGs include *Everquest*, *EVE Online* and *Lineage*. Communication in MMOGs is examined in study 1 (chapter 4).

Social worlds also simulate large, often fictional spaces, but are not considered by their users to be games (though mini-games can be set up within a social world). There is no formal competition and nothing to ‘win’. There are fewer rules constraining users, and these systems are used for a broader range of activities including social interaction, discussion, role-play and content creation. The most popular social world currently is *Second Life*, which has about a million users. Other prominent social worlds have included *There*, *Habbo Hotel* and *Active Worlds*. I examined social worlds in studies three and four.

Thus while early CVEs were imagined as systems for computer-supported collaborative work, they have succeeded in the form of recreational technologies. Yet this evolution has recently gone full circle, with interest growing in the appropriation of VWs for instrumental tasks such as online meetings and distance education. Linden Lab, the vendor of *Second Life*, actively promotes its system to business and education users, and other systems designed specifically to support collaborative work have been developed.

The timeline in figure 2 illustrates a brief history of virtual environments to date.

![Figure 2: Timeline of virtual environment technologies](image)

A fundamental problem in the design of desktop virtual worlds is how they can provide a usable interface to a simulated 3d space despite using 2d hardware such as a mouse (Bowman et al., 2001). Broadly speaking this lack of a third input dimension has been addressed in two
ways. One has been to devise paradigms that superimpose a third dimension onto existing 2d interfaces. For example, a modifier key can define some mouse movement as being in the 'z' direction. The other approach is to invent new I/O devices that are explicitly three-dimensional, such as the 3d mouse and the stereoscopic display. But virtual worlds have achieved mass uptake by ‘making do’ with standard hardware. Only very recently have true 3d devices such as the Wii Remote and Xbox Kinect become available to average consumers.

I am less concerned with how VW users communicate with their computer, and more with how they communicate with each other. (Note that my topic is distinct from, say, the use of speech for inputting system commands.) To use the terminology of Preece and Maloney-Krichmar (2003), I am concerned more with the sociability of virtual environments than with their usability. Despite this emphasis, part of the challenge of designing successful communication tools for use in VWs stems from their spatiality.

As the world grapples with economic and ecological problems, Internet-enabled virtual worlds offer the promise of an inexpensive platform for collaboration among geographically-dispersed users. For example in 2008 Imperial College London and Nature Publishing Group co-hosted a conference in Second Life with the explicit aim of reducing the carbon footprint associated with long distance travel to international events\(^2\). Likewise many universities are investigating the use of VWs as teaching spaces for distance education (e.g. Gregory et al., 2011), with aims that include easing the travel burdens of students and staff.

High expectations have been held for CVEs since their earliest incarnations. At times it has been predicted that they would become a dominant platform for online interaction (e.g. Rheingold, 1991; Gartner, 2007). However it would be fair to say that no implementation has thus far lived up to these expectations (Salomon, 2009). Commercial success is mostly limited to entertainment applications, and CVEs are less often put to serious use outside the research lab (Schroeder, 2010). Even the large MMOGs which boast millions of users do not approach the broad popularity of, say, email or social networking sites, whose users number in the billions. It is an open question why more people have adopted web-based social networking than have adopted virtual worlds. Researchers have offered critiques of existing implementations, as well as of over-enthusiastic expectations for 3d (e.g. Harrison and Dourish, 1996). One thing is certain: merely including three-dimensionality within a communication system has not guaranteed the “natural articulation of collaboration” (Benford, et al., 1994) that has been envisaged for virtual environments.

My experience studying and using virtual worlds has led me to conclude that, in order to estimate the aptness of VWs to a particular communication scenario, one needs to understand them: (a) as multi-media technologies, of which three-dimensional space is one of the media, and (b) as one, perhaps extreme, example of a broad project to build communication technologies of maximum ‘richness’ in the sense of Daft and Lengel (1986).

To describe CVEs as multimedia technologies is to emphasize that mechanisms for simulation of space, embodiment of users as avatars in the space, and linguistic communication between users, while they are frequently found together, are independent and need not all be implemented together within a given technology.

Some discourse about virtual worlds has implied that a VW’s purpose is to provide rich communication, even if the academic term ‘media richness’ is not explicitly used. For example Wasko et al. (2011) explain that:

According to theories of media richness, 3D environments are objectively rich because there is synchronous contact; the visual stimuli, objects, and environmental designs offer a variety of social cues; and communication occurs through multiple channels, including audio, visual, and text. (p. 648)

A medium’s richness has been understood as the being extent to which it imitates face-to-face conversation by conveying not only linguistic messages but also information about the people conversing (Daft and Lengel, 1986). This view of communication sees face-to-face (f2f) conversation as the ‘gold standard’ to which mediated communication is to be compared. F2f communication includes not only speech but facial expressions and body language conveyed through the modality of vision. Thus an explicit goal of much VW design has been to simultaneously transmit, along with speech or text, simulated bodily orientations and gestures through the users’ avatars.

Implicit in the media-richness project is the belief that richer communication is better communication. This means that a CVE can fail in two ways. It can fail to be rich because of technical limitations, or it can fail because, in some situations, richer does not mean better. While the former is sometimes offered as an explanation for the limited success of CVEs, so that the solution must be to seek even greater fidelity of representation, in this thesis I argue that the latter is often the better explanation. In other words, VW users, like people in general, do not want maximally-rich communication media.

For example, VW users do not usually want their avatars to resemble them exactly (Ducheneaut et al., 2009), and my studies showed that some of them choose to communicate by text, even when voice is available, in order to prevent audio cues about their identity and physical context from being transmitted. Media richness theory has been much critiqued (see
chapter 2), yet this critique appears not to have permeated discourse on virtual world design. Clarifying this situation is one of the aims of this thesis, and is discussed in detail in chapter 8.

1.2 Classification and terminology

1.2.1 Virtual space

This thesis is concerned with computer systems that ‘locate’ their users within a simulated three-dimensional space and allow them to ‘move’ within it. I place ‘locate’ and ‘move’ within quotes because, of course, the positions at which users are ‘located’ are points in a simulated space. The users of a virtual environment do not need to be at any particular physical location while moving within virtual space: a user’s physical and virtual locations are, in existing VWs, independent of each other. In this thesis I address both the physical and virtual contexts of use, so I need to be clear that these are distinct. The user’s physical context is the space immediately around the computer they are using, and may include other people who can see or hear what the user is doing. (Modern high-powered laptops and mobile Internet connections mean that a VW user’s physical context can be almost anywhere.) The user’s virtual context is the virtual space immediately around their avatar. It might contain the avatars of other users who because of avatar proximity can see the first user’s avatar, receive their text chat, and through a voice channel possibly hear what is happening in their physical context.

Virtual worlds are computer-generated Euclidean spaces with three dimensions (‘3-spaces’), because that is the dimensionality of the space that we experience in our everyday dealings with the physical world\(^3\). In three-dimensional space the position of any object can be described by three coordinates, which represent the distance from an agreed origin-point to the object along three orthogonal axes. The orientation of an object (the direction it faces) can be described by another ordered triple of angles relative to the agreed axes. Accordingly, translations (changes in location) and rotations (changes in orientation) can be expressed as three-dimensional vectors, and distances between points can be calculated straightforwardly by

---

\(^3\) In fact a VW can be described as a Cartesian space because it has a coordinate system. Coordinates are necessary for the software to function, so distinguishing between Euclidean and Cartesian is unimportant. The important distinction is between these and non-Euclidean spaces.
the Pythagorean formula. In a Euclidean space, the distance between two objects, along with their orientations, is part of what determines whether they can affect each other and thus whether they are relevant to each other (Benford et al, 1994).

Alternative models have been proposed to describe the space of the physical universe. For example, relativity utilizes a non-Euclidean space, and string theories propose higher dimensionalities that only become significant at tiny distance scales. These do not affect physics at the scale of human experience, and no virtual environment has attempted to represent them. (Dix et al., 2000, discussed the non-Euclidean spatialities of modern physics and their potential for application within other types of virtual spaces, such as mobile-device networks and the World-Wide Web.) The fact that virtual environments are Euclidean is not discussed further in this thesis: I simply refer to them as ‘spatial’, ‘three-dimensional’ or ‘3d’, and speak of them as having the property of ‘spatiality’.

Numerical coordinates can be used explicitly by *Second Life* users in order to directly manipulate locations, such as when building and scripting objects or ‘teleporting’ around the virtual world. By contrast, in MMOGs and game-worlds, the 3d nature of the space is often not made numerically explicit to users, and MMOG users cannot directly manipulate either their own location or those of objects in the game-world. However from the software developers’ point of view, all these spaces are explicitly three-dimensional.

There are different expressions in use for the properties and phenomena of virtual environments. However while there is not complete agreement on terms, in this section I will establish a nomenclature to be used consistently throughout the thesis.

Since the defining feature of virtual environments is their simulation of 3d space, I should first differentiate them from:

- networked multimedia systems such as hypertextual spaces (e.g. the world-wide web),
- text-based ‘MUDs’, which describe a pseudo-spatial environment using words rather than graphics,
- systems that represent a space that is navigable but two-dimensional, and
- cinema-like systems that display 3d scenes which are not navigable (Nitsche, 2008).

Some videogames use a so-called “2.5d” projection. The environment is two-dimensional but is displayed as if looking from above, to give the impression of three dimensions. One 2.5d
world, *Ultima Online*, is considered by many to be the first MMOG. Others such as *Club Penguin* and *Habbo* are popular among young people. While 2.5d systems may certainly be experienced as ‘worlds’ by their users, I have not included them among my cases. They do not afford the degree of spatial interaction possible in 3d systems.

The word ‘virtual’ has several usages. Informally it has been applied to a wide range of Internet-based technologies, so that websites offering e-commerce have been called ‘virtual stores’ and web-based learning management systems are sometimes called ‘virtual classrooms’. That usage – by which the word ‘virtual’ means ‘web-based’ – has waned in favour of using the term only to refer to spatial simulations, and it is these with which I am concerned.

The relationship between virtual worlds and other Internet-based technologies is illustrated in figure 3.

*Figure 3: Relationship between CVEs and other technologies*
1.2.2 Types of collaborative virtual environments

There are several types of CVE, including:

- fully-immersive VR systems running on specialized hardware
- multi-player 3d videogames, of which the ‘first-person shooter’ (FPS) is a prominent genre
- virtual worlds.

Related technologies include video-conferencing systems, and immersive video projections such as the CAVE (Johnson and Leigh, 2001). Video-conferencing does not generate space: rather it projects an image of one real-world space into another (though Schroeder, 2010, predicts that non-PC-based CVEs and video-conferencing systems will converge somewhat). Immersive projection systems mix physical objects with virtual space, using specialized, expensive hardware that is not widely available. These are not within the scope of my thesis: Nor are specialized hardware platforms such as location-based entertainment centres and theme park rides (Badiquet et al., 2002). I have limited my research to the more widely-used, if less visually immersive, PC-based virtual worlds.

When analysing virtual worlds it is necessary not only to classify technologies but to differentiate the kind of uses to which they are put. Broadly speaking, researchers differentiate two approaches to the use of VWs.

- Leisure, entertainment, socializing etc (I call this “recreational”), and
- Utilitarian or work-related use (I follow Schroeder, 2010, and call this “instrumental”).

1.2.3 ‘First person shooter’ videogames

First-person shooter (FPS) videogames are not normally classified as VWs because their space is not persistent and they do not sufficiently many users in one environment. However in my second study I used a voice system designed for team-based FPSs, as no voice system of this type had been built for true virtual worlds.

One of the most popular FPSs is Counterstrike. One of the earliest, Doom, displayed scenes as though the user were looking through the eyes of their character: hence the term ‘first-person’. Recently most FPSs represent the user as an avatar in the game-world, situated just in front of
the retinal plane so that the user views the scene from behind their character’s head. Some call this a ‘third person’ view, however ‘first-person shooter’ has prevailed as the name of the genre.

When FPSs are multi-player it becomes important to represent users visually, because a significant subset of the environment that a player is attempting to negotiate consists of other players. The avatars of fellow players may be targets or foci of action. Like the pieces in chess, the positions of players in virtual space is a critical part of game state in a FPS.

FPSs are described in more detail in section 2.2.3.

1.2.4 MUDs: text-based worlds

Multi-user-dungeons or MUDs are a kind of online world in which the environment is described for the user in text rather than being graphically illustrated. Much of the thematic content of current MMOGs is inherited from MUDs, and some MUDs are still in use.

A MUD’s space is not truly 3d (they can be non-Euclidean networks of rooms – see Aarseth, 2008), and is not represented visually, so MUDs do not fit most definitions of ‘virtual environment’. I did not examine MUDs in this thesis; however they are of historical importance and research into use of MUDs is discussed in section 2.2.2.

Because MUDs are sometimes described as “text-based virtual worlds”, I need to clarify that when I discuss ‘text communication’ I am not referring to the method by which a MUD displays its environment to users, but to the typed messages that people use to communicate within graphical virtual worlds.

1.2.5 Using virtual worlds to research offline behaviour

While much research has sought to understand phenomena within virtual worlds (see section 2.2), some authors have proposed that research into virtual worlds can be used as a basis to study human behaviour in general, and that the VW context may be more convenient than the offline.

For example, Castronova (2008) tested economic theory within virtual worlds to support their use as models of offline economic activity. Bainbridge (2007, 2010) argued that the study of Second Life and World of Warcraft can offer a glimpse of how people may interact in future.
A taxonomy of the systems relevant to my thesis is presented in figure 4. The red boxes represent genres that were not covered in my research.

![Figure 4: Taxonomy of virtual environments](image)

### 1.2.6 Virtual and physical space

It is common in discourse about virtual worlds to compare virtual and physical space. Several terms are used for the physical world and there is some debate over which is most suitable. Nitsche (2008) refers to a videogame player’s physical context as ‘play space’, and the collection of the play spaces of several collaborating or competing players as ‘social space’. Second Life users typically refer to the physical world using the term ‘real life’ (‘RL’), or less commonly ‘first life’ (‘FL’) or the ‘full bodied’ world.

Boellstorff (2008) argues that the use of the term ‘real life’ is misleading, since all human culture acts as a kind of virtual world in which social life is enacted. He recommends the term ‘actual world’, though this has not been widely adopted (see Golub, 2010, for a defence of this terminology).

This thesis is concerned with the mechanics of communicating within virtual space. Therefore I chose the term ‘physical world’ to describe the everyday world. This term emphasizes that communication in the physical world is subject to physical forces, which must be simulated in virtual worlds in order for communication to take place. In the physical world, voice conversations are usually propagated by sound waves or electromagnetic signals. Virtual worlds must explicitly implement a mechanics of signal propagation in software. These
mechanics can have arbitrary design and need not (and usually do not) mimic precisely the mechanics of the physical world.

To distinguish physical and virtual space is not to imply that virtual space is a separate universe untouched by real-world actions or social institutions. The idea that virtual worlds are walled off by a ‘magic circle’ has been dismissed by Lehdonvirta (2010), supporting earlier arguments by Taylor (2006) and others. One of my findings is that the modality of voice tends to breach whatever boundary does exist between virtual and physical worlds.

### 1.3 Theme: Communication in virtual worlds

Virtual worlds offer several means for user-to-user communication:

- the simulated space, which is visualized to each user and can support indexical utterances such as “this thing” and “that location”,
- user embodiment as avatars, which affords a simple body-language of proxemics, posture and gesture,
- linguistic communication via text and/or speech,
- sound, triggered by the user actions on the environment,
- and in a few systems, haptic or other sense modalities.

Because VWs locate users in a space, designers have the option of taking the sender’s and receivers’ virtual locations into account when implementing mechanisms for the transmission of signals. For example, spoken or typed messages might be sent only to users whose avatars are ‘near’ (in virtual space) the sender’s avatar, simulating the transmission of sound in air and enabling “a natural intuition about mutual audibility” (Smith et al., 2001). Conventionally this is called proximity, spatial or local chat.

Alternatively the VW might ignore avatar locations and simply transmit messages between users, wherever they are. In effect, this simulates the use of telecommunication devices.

Thus I have suggested that there are in effect two metaphors currently used to design voice transmission in a virtual world: sound in air, and radio (see Wadley et al, 2005 in appendix C).

The configurations typically available in VWs are illustrated in table 1, along with the terms commonly used to describe them.
<table>
<thead>
<tr>
<th>Metaphor</th>
<th>Text modality</th>
<th>Voice modality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sound-in-air metaphor</td>
<td>‘local chat’, ‘vicinity chat’</td>
<td>‘spatial voice’, ‘proximity voice’</td>
</tr>
<tr>
<td>Radio metaphor</td>
<td>‘instant message’, ‘guild chat’, ‘raid chat’</td>
<td>‘group voice’, ‘one-on-one’</td>
</tr>
</tbody>
</table>

Table 1: Metaphors for propagation of voice and text messages

Whichever metaphor is used, the rules of message propagation must be explicitly designed by the system developers. These rules constitute a simulated medium through which messages are propagated. Therefore while networked virtual environments are themselves a kind of telecommunication medium, they are unique in that they simulate a 3d space, locate users within it, and then provide them with one or more *virtual telecommunication media* through which they can communicate within virtual space.

This complexity relative to other communication technologies raises a number of issues:

- How should virtual media be configured? Existing implementations are usually simulations of sound or radio, but they could take arbitrary forms.

- What are the relative efficacies of differently-configured virtual media when used by different kinds of users in different situations?

- How can users manage the multiple communicative contexts that VW users experience?

These questions are all addressed within this thesis.

In the next section I describe the problem my work has attempted to address, after which I offer a précis of relevant prior research. I then identify a gap in this literature, present my research questions, give an overview of my research program, and précis my results and the conclusions I have drawn.
1.4 Problem

As discussed above, many virtual worlds now include mechanisms that allow users to speak to each other. Yet as the literature reviewed in the next chapter shows, there has been little systematic investigation of the influence that the voice medium has had on the experience of virtual world users. A thorough study has not previously been made of what advantages voice brings, or conversely which conditions, if any, render voice unsuitable. It has not been established who finds voice unsuitable, nor how this might be addressed by designers.

The influence of voice in virtual worlds is of theoretical interest because these systems are considered by some to be arenas *par excellence* for the kind of identity-play that in the past was held by many to occur within the Internet more widely. If VW users are exploiting ‘lean’ text chat and customizable avatars to enact fictitious personas in imaginary worlds, then voice should either make their role-play more difficult, or force us to reconsider what role-play is. The study of voice in virtual worlds can therefore inform our understanding of mediated communication more broadly.

The study of voice in virtual worlds has practical importance as well, because without understanding the experience of users who communicating by voice, virtual world developers run the risk of failing to successfully implement these large, expensive systems.

Some voice implementations have already been rejected by users; thus it would appear that such failures of understanding have already occurred. For example, the voice system added to *World of Warcraft* by its developers has been rejected in favour of third-party voice systems (Street, 2011). The introduction of voice to *Second Life*, too, met with controversy (chapter 7). Many users reacted angrily to Linden Lab’s announcement of their intention to add voice. Some not only refused to use voice but even threatened to quit the VW altogether if voice was introduced (Boellstorff, 2008). This indicates that not only the voice sub-system but the virtual world itself is at risk if implementation goes awry.

Finally, VW communication mechanisms need to be usable by a large, heterogeneous user base, and the design space for voice is potentially broad and has only been superficially investigated. Therefore an understanding of how different configurations are received by different users conducting activities in different contexts will be of value.
1.5 Prior work

In reviewing prior research I have focused on two fields: the study of virtual worlds and the study of communication.

Research into virtual worlds has focused on a number of issues. Authors such as Turkle (1995) and Kendall (2002) were interested in the ability that MUDs afforded users to engage in presentation-management and role-play. Bartle conducted pioneering work on MUDs, proposing a framework of four user types: achievers, socializers, explorers and killers. Yee (2006) developed this for MMOG players, surveying large numbers in order to explore their demographics and motivations.

There has been research on how MMOG users behave while playing (Moore et al., 2007), interact with other users (e.g. Seay et al., 2004), and use, customize, and relate to their avatars (e.g. Ducheneaut et al., 2004, 2009). Ethnographers have studied the culture of MMOGs (Taylor, 2006; Nardi and Harris, 2006; Ducheneaut et al., 2006; Bainbridge, 2010; Nardi, 2010; Golub, 2010), and social worlds (Taylor, 1999; Boellstorff, 2008).

Communication research has compared media and examined how these influence people’s choice of medium and the communication they subsequently carry out. The ‘social presence’ (Short et al, 1976) and ‘media richness’ (Daft and Lengel, 1986) theories proposed that the richer a medium - the more it projects the social presence of users - the more effectively it should substitute for face-to-face interaction. Subsequent research has proposed other influences on media choice and use. The social-influence (Fulk et al., 1990) and critical-mass (Markus, 1990) frameworks proposed that an individual’s choice of medium is influenced less by the medium’s properties than by the choices of the individual’s collaborators. Walther (1996) argued that the reduced social presence of text-based computer-mediated communication provided an increased opportunity for management of self-presentation, which enabled hyperpersonal interaction: a more intense relationship than would be expected to occur offline. Research has shown that people wishing to engage in deception or ‘impression management’ (Carlson et al., 2004), or who are shy (Stritzke et al., 2004), might prefer a medium with low social presence such as text.

A more extensive survey of relevant prior research is presented in chapter 2.
1.6 Gap

Voice is a relative newcomer to virtual worlds, and only limited research has compared modalities in this context. For example, Sallnas (2002) compared decision-making by VW users equipped with text, voice, or a video link. Nilsson et al. (2002) supplemented *Active Worlds* with a shared-audio system and studied workplace meetings held in the virtual world. Gibbs et al. (2004) and Halloran et al. (2004) studied the use of voice by FPS players, finding that there were benefits, particularly concerning sociability and freeing up of hands to use game controls, but that these were situation-dependent. Williams et al. (2007) found that *World of Warcraft* teams who used both voice and text to communicate over a period of months, liked and trusted each other more, and became happier and less lonely, than those who communicated only by text.

When my research began in 2006 it was uncertain what influence voice had on the experience of VW users, why it tended to provoke extreme like and dislike, how existing implementations might be better configured, how voice could interact with the spatiality of a VW or what new forms of interaction it might enable.

1.7 Research Question

Given this gap in knowledge, I chose as my research question:

RQ: *How does voice influence the user experience of virtual worlds?*

To answer this question I conducted four case studies designed to address the range of virtual worlds, situations and users discussed in section 1.2. My studies are outlined in the next section, and described in detail in chapters 4 through 7.

1.8 Approach

To detect patterns across the diversity of technologies, situations and usages that exist I conducted a series of case studies (Cavaye, 1996; Yin, 2003) of communication in virtual
environments. I employed a range of methods to study the influence of voice on user experience in these various scenarios. My methods are described in detail in chapter 3.

Understanding subjective experiences requires “accessing the meanings participants assign to them” (Orlikowski and Baroudi 1991). This implies an interpretive approach which acknowledges that users interpret their own experiences and that the researcher in turn interprets what users report (Neuman, 2011). I gathered subjective data via individual and group interviews and diaries, but triangulated these against the results of observation, participant research, and quantitative analysis. Use occurred in natural settings such as homes and workplaces, except in one study which took place in a lab but was designed to be as naturalistic as possible (chapter 6).

My approach to case selection, data gathering and analysis was informed by grounded theory (Glaser and Strauss, 1967). The principle of theoretical sampling guided my choice of cases, particularly when moving from the first to the second pair of studies. My first two studies examined game worlds: after analysing these it was apparent that the appeal of voice was partly due its support for real-time coordination of teams during fast-paced action. To test this I chose non-game worlds for the second pair of studies.

Participants were recruited from diverse populations (within the constraint of their needing to be available for interviews), and had a range of experience with the technologies in question. The technology used in all cases were commercially available, except for one voice system which was under development during the study and has since been commercially released.

**Study 1: Voice in massively multiplayer online games**

My first study examined voice communication among people playing MMOGs. I arranged for three groups to play under different circumstances over a period of two months. All participants had some experience with MMOGs, and some had already used voice products in online games.

At the time of the study, the first MMOG with integrated voice facilities had recently been released. Two groups played this, while a third group played older MMOGs using third-party voice products. They played in their own homes under normal playing conditions, and kept diaries in which they recorded experiences of the use of voice in the game. Half way through the study the participants were interviewed individually. At the end of the study they participated in focus-group interviews.
Participants were asked questions about whether they preferred voice or text, whether either modality was better suited to particular scenarios, and whether there were aspects of the voice interface they would like to see changed. The focus groups considered how they might use existing or imagined voice systems to deal with fictitious gameplay scenarios.

**Study 2: Spatial voice in a team game**

My second study examined the influence of spatially-propagated voice on users’ experience of a team combat game. At the time of the study, existing voice systems utilized a radio metaphor which assigned a channel to each team and allowed all team members to hear each other equally. I examined use of an experimental system in which the ability of one user to hear another was based on the proximity of their avatars in virtual space. Thus communication was explicitly integrated with the spatiality of the system.

Spatial voice was provided to a group of co-workers to use in a weekly game session. Participants were observed and asked to keep a diary in which they recorded their experiences. They used spatial voice for several months before taking part in a focus group.

**Study 3: Collaboration around objects**

My third study again addressed spatiality and communication. But instead of the ‘macro’ spatiality involved in coordinating a moving team, study 3 focused on the ‘micro’ level of collaboration around objects. Here the key problem is how to achieve mutual understanding of reference to locations and things.

My design combined the quasi-experimental approach of Hindmarsh et al. (2001) with methods from Kraut et al. (2002). I observed and interviewed small groups using voice to coordinate a building task in Second Life. I recorded their screen output and conversation for analysis. I focused on their verbal references to objects, and use of Second Life’s ‘virtual camera’. I discussed with participants the problems they faced and how they solved them. I checked my observations by discussing themes that emerged with expert users discovered in-world\(^4\) and in the SL forum.

\(^4\) The expression “in-world” is used by Second Life users to denote events and places in the virtual world.
Study 4: Voice in the ‘social’ virtual world Second Life

My final study examined the overall influence of voice on the social world Second Life. I interviewed users, convened an in-world discussion, analysed forums and blogs, and conducted participant research into both recreational and instrumental uses of SL. The participants used Second Life for a diverse range of activities that included socializing, teaching, business and art. This very broad range of use scenarios allowed me to compare findings with the first three studies and draw general conclusions about voice in virtual worlds.

The four cases are compared in table 2:

<table>
<thead>
<tr>
<th>Type of VW</th>
<th>Voice propagation</th>
<th>Research methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 1</td>
<td>Massively multiplayer online game</td>
<td>Radio</td>
</tr>
<tr>
<td>Study 2</td>
<td>First-person shooter</td>
<td>Spatial + radio</td>
</tr>
<tr>
<td>Study 3</td>
<td>Social world</td>
<td>Spatial</td>
</tr>
<tr>
<td>Study 4</td>
<td>Social world</td>
<td>Spatial</td>
</tr>
</tbody>
</table>

Table 2: Overview of my four case studies

1.9 Outcomes and contributions

Study 1 showed that voice provides significant advantages for team coordination, and can make the MMOG user experience more sociable. However it increases the emotional intensity of user-to-user communication and can intensify the effects of griefing. Voice makes it difficult for users to maintain a ‘fictional social presence’, and to simultaneously manage physical and virtual contexts.
Study 2 found that while spatial propagation is a restriction over radio, it reduces channel-clutter and acts as a ‘filter for relevance’ by conveying to users only the utterances of fellow users who are nearby in virtual space. Making members of competing teams audible to each other creates new opportunities for tactics.

Study 3 showed that achieving successful reference to objects and locations remains difficult in a modern virtual world. Users engaged in object-focused work mark locations not by pointing but by using their avatars as ‘cursors’. At times they simply ‘parked’ their avatars. This calls into question the necessity for embodiment in avatars, and suggests that virtual worlds instead allow users to represent their focus-of-attention directly on objects.

Study 4 highlighted the diversity of purposes and contexts for virtual world use. Users have difficulty managing multiple identities and conversations in both physical and virtual contexts. Choice of communication medium is of critical importance to VW users. The criteria by which they judge media cannot be reduced to a simple, single concern.

1.10 Conclusion

My research has allowed me to challenge several ideas regarding mediated communication in virtual environments:

- Voice is not straightforwardly superior to text in virtual worlds, any more than it is in physical environments. In fact it is problematic in many situations.
- Virtual worlds are not cut off from physical reality. The fact that users are embedded in both physical and virtual contexts has an important influence on their choice and use of communication modality.
- Avatars may not be essential for conveying social presence or supporting deictic speech.

In this chapter I provided an overview of virtual worlds research, illustrating the gap I hoped to fill, and described my research approach and findings. In the next chapter I review in greater detail prior research into virtual worlds, mediated communication and user experience.
Chapter 2

2. Review of related research

2.1 Introduction

In the previous chapter I presented an overview of my thesis topic: the influence of voice communication on the user experience of virtual worlds. In this chapter I review prior research from fields relevant to the topic.

To identify relevant literature and organize this chapter, I have first isolated the components of my research topic. I am interested in the intersection of three fields:

1. virtual worlds,
2. mediated communication, and
3. user experience.

I first highlight important phenomena from each of the fields, in sections 2.2, 2.3 and 2.4 respectively. Then in section 2.5 I consider how these phenomena might interact at the intersection of the fields. This allows me to identify a gap in knowledge and deduce my research question.

2.2 Virtual worlds

In section 1.2 I identified virtual worlds as a subset of the category of technologies known as collaborative virtual environments (CVEs). These multi-user technologies simulate a three-dimensional space containing virtual objects, some of which typically are representations of the users themselves (Aarseth, 2008; Nitsche, 2008). The users of a CVE can interact and communicate with each other, and may cooperate or compete to achieve goals.

Data about the contents of a CVE are stored on a server, to which users attach via client software. At every given moment the client renders for the user a visualization of the space and
its contents: this must be done from a point in the space, so that the user can be considered to be ‘situated’ at that point. ‘Moving’ in the virtual space essentially means changing the location of this point, or the direction in which one looks through it.

This location and orientation in virtual space is typically made visible to other users of the space by representing the user graphically as a avatar. Avatars may be human-shaped but this varies according to the theme of the VW. This mutually-visible, embodied representation is designed to allow users to simulate offline social interactions that involve the negotiation of space, such as conversation groups and team-based combat, to choose two common examples.

Avatars afford their owners considerable plasticity of appearance, and therefore pseudonymity or anonymity (Bente et al., 2008). Part of the appeal of using an avatar is the control it gives a user over their social presence within the system. However this depends on the system and the type of use people make of it.

Avatars also exhibit behaviour. Users of PC-based virtual worlds control their avatars’ movements and gestures through deliberate control of the mouse and keyboard. On the other hand some advanced, immersive CVEs running on specialized hardware can detect user gestures and even facial expressions. Hence in these “an avatar is the model that is rendered on the fly to reflect the user’s behaviour” (Bailensen et al., 2006). Opinions differ as to whether either fidelity to the user’s movements or plasticity is the more desirable. Schroeder (2010) felt that the desirability of fidelity depends on whether a CVE is to be used for instrumental or leisure use, and regarded fidelity and plasticity to be suited to immersive and PC-based systems respectively (p. 22).

Many of the issues relevant to current virtual worlds were first identified in early CVE research, and I review this first.

### 2.2.1 Early collaborative virtual environments

CVEs evolved out of experiments with interactive 3d graphics (e.g. Fisher et al., 1986), vehicle simulations such as pilot trainers, and systems for the remote operation of machines in oceanic, Antarctic and extra-terrestrial exploration (Ellis, 1994).

The first virtual reality systems simulated an artificial environment for one user, and were characterized by the use of a head-mounted display (HMD), which displayed the environment in such a way that the user’s head movements caused their view to change, as happens when viewing the physical environment. Some VW systems supported the use of data-gloves for
haptic interaction and hand gestures (Snowdon et al., 2001). Some users of head-mounted displays encountered problems with disorientation and nausea (Ellis, 1994). Other research issues included interactivity, visual realism, speed of performance, and immersion. Ellis (1994) suggested applications of VR for surgery, robot control, data visualization, and entertainment.

VR systems were typically used by one person at a time. The collaborative system of Codella et al. (1992) supported two users who were represented to each other simply as hands: interaction was limited to joint manipulation of objects.

Subsequent CVEs have typically represented users as humanoid avatars. Researchers studying avatar use have been interested in whether these assist in providing “peripheral awareness, informal meetings and learning by watching” (Jaa-Aro and Snowdon, 2001).

A number of graphical multi-user environments were constructed for research purposes during the 1990s: prominent examples include MASSIVE and DIVE.

### 2.2.2 MUDs

Graphical virtual worlds were preceded by MUDs, and inherited themes and user culture from them. MUDs (multi-user dungeons) are networked games that simulate a space within which users move and communicate; however the space is non-Euclidean and is portrayed with words rather than with graphics (Bartle, 2008).

The textual medium afforded pseudonymity and a significant capacity for presentation-management, allowing MUD users to present as fictitious characters and to play roles:

Users are not required to ever present their RL identities. [They] may therefore perform actions they might not in a face-to-face encounter. …. interlocutors decide for themselves the degree to which they wish to reveal facets of their identities and are not limited to enacting their RL roles. [They] are more aware of the presentation of the self and have more control over whom others perceive them to be. …. The anonymity of computer-mediated communication in [MUDs] supports playful experimentation with one’s identity, and ultimately mastery over presenting the self to others in a virtual environment” (Raybourne, 2001).

Bartle, a pioneering developer of MUDs, argued (1996) that the primary reason for playing a MUD was to role-play in such a fashion, in order to develop one’s own (real life) character. Bartle recognized different types of players with different styles of use: achievers, socializers, explorers and killers, and argued that individual players enacted a “career” during which they shifted from one style to another.

Schiano (1999) studied the large-scale MUD LambdaMOO, finding that users emphasized social interaction over navigation of the space.
Kendall’s (2002) ethnography of the BlueSky MUD used the metaphor of a ‘virtual pub’ whose pseudonymous patrons maintained an informal sociality:

As usual around lunchtime, the bar is crowded. A few people sit singly at tables, but most sit in small groups, often milling around from table to table to chat with others. As in many such local bars and pubs, most of the regulars here are male. Many of them work for a handful of computer companies in a nearby high-tech industry enclave. The atmosphere is loud, casual, and clubby, even raucous. Everybody knows each other too well here to expect privacy at any of the tables. (p. 2)

MUD users multi-tasked as they divided their attention between physical and (possibly several) virtual spaces. Many utilized more than one character or MUD, engaging in multiple ‘presentations of self’. However Kendall, citing Goffman (1959), pointed out that:

“people also engage in different presentations of self to different audiences in other arenas of everyday life and did so before online forums existed (p. 9”).

Kendall found that the MUD was a male space in which “people enact and negotiate masculine identities” (specifically, the ‘computer nerd’ style of masculinity). She cautioned against earlier utopian views of the Internet as a system offering infinite possibility for identity construction (e.g. Turkle, 1995), arguing instead that MUD users bring their offline gender, race and class identities into the virtual world.

Despite the MUDs’ text-only representation, users were able to achieve rich, long-term interaction. Churchill and Bly (1998) studied a group of co-workers who used a MUD to communicate within a workplace over a period of three years:

Our studies indicate that visually oriented media richness is not a prerequisite for the creation of sufficient social co-presence for maintaining collaborative relationships. Such social co-presence seems to reside in shared goals and understandings which derive from conversations around a common focus. (Churchill and Bly, 1998)

The low fidelity of textual representation gave MUDs an advantage over video- and audio-based media-space technologies with regard to privacy. Churchill and Bly found that co-workers valued being able to control the information about themselves and their surroundings that was transmitted into the shared medium:

The fact that one actually cannot see or hear what is really going on in others’ offices offers a significant advantage to the MUD. The impression a participant wants to share with others is determined by that participant and not by the physical or audible context. The MUD offers cognitive co-presence but not physical or visual co-presence. (Churchill and Bly, 1998)

The authors recognized that this conclusion “flies in the face of many theories of media richness”.

2.2.3 Networked FPS games

A step in the evolution of modern virtual worlds was the development of LAN-enabled first-person shooters (Castronova, 2005). These combat simulators run on PCs connected via a local network, enabling cooperative play in a graphical 3d environment. Participants in a LAN game are physically co-present, and thus visible and audible to each other, allowing them to communicate using the speech and gesture normally employed in face-to-face communication.

Taylor and Witkowski (2010) described a large LAN party held in Scandinavia in which 14,000 players participated in a range of multiplayer games and other activities:

“These realtime face to face events bring together people, and their machines, for several days of intensive interaction and play. LAN parties can involve everything from file sharing and demos to game playing and other activities. They are often a mix of people who already know each other to strangers coming together to meet for the first time. They can range in size from a handful of people dragging their machines over to someone’s apartment for an evening to large scale, highly coordinated events with tens of thousands of participants.” (2010, p. 1).

Aarsand and Aronsson (2009) analysed talk around a home game console to demonstrate that when co-located people play a game together, or watch while others play, there is constant verbal communication and in particular frequent “response cries”. Response cries include the exclamations and mild profanities which these authors described as “public displays of emotions in the form of vocalizations such as self-talk, imprecations, and audible surprise” (p. 1557). These verbalizations are not only communicative but performative, and occur in response to the activities and themes of games being played: “singing along, making sound effects, producing response cries and animations, all communicative actions that can be seen to form a type of action aesthetic” (p. 1567). This study also highlighted the different roles that those co-located perform such as ‘player’, ‘spectator’ and bystander’.

By contrast, the online context forces interaction to be mediated and therefore to depend on which media are available. Connecting network games to the Internet opened them to large numbers of unknown and unsupervised users and contexts. Reeves et al. (2009) studied the most popular Internet-based FPS Counterstrike, providing a detailed analysis of the skill required to negotiated this fast-paced, team combat game. Fast reactions are required, as are complex sequences of actions including moving through terrain, taking cover, identifying and shooting enemies, and coordinating with team-mates. Counterstrike supports team coordination through a voice channel provided in the game software, and this contributes to mutual awareness by allowing players to
combine a close eye on the current proceedings with a complex and developed sense of what others—friend and foe—are doing in the environment. (Reeves et al., 2009, p. 25)

A key difference between LAN and online gaming is that collaborators and competitors connecting via the Internet cannot see or hear each other, and frequently do not know each other. Teams connecting via the Internet rely upon the tools of computer-supported cooperative work such as synchronous, computer-mediated, text- or voice-based communication, and asynchronous tools such as blogs and forums. It is with online communication tools such as these that this thesis is concerned, and they are described in section 2.3.

Internet-based gaming is now the more common form of multi-player gaming and connects millions of players worldwide via personal computers and dedicated game consoles. The geographical dispersion inherent in Internet use has had such an influence on the user-experience of multiplayer gaming that Taylor and Witkowski (2010) reported being struck by the co-presence, spectatorship and bursting of the “personal space bubble” that occurs at LAN parties.

Though the virtual environments of networked FPSs were not persistent, long-term associations of users (“clans”) developed, which historically were the precursors of MMOG guilds. MMOGs, in which large numbers of players interact in a persistent virtual environment, can be considered a combination of the technology of Internet-based FPS with the game-play, culture and themed settings of MUDs.

### 2.2.4 Massively Multiplayer Online Games

The first graphical MMOG, *Meridian 59*, was released in 1996. The first wave of MMOGs that attracted player numbers in the hundreds of thousands included *Anarchy Online* and *Everquest*. One of the most important and popular MMOGs to date, *World of Warcraft*, was released in 2004 and currently boasts 11 million users.

Academic interest in MMOGs has addressed several issues. Yee (2006) surveyed 30,000 MMOG players between 2000 and 2003 to explore their demographics and motivations and the impact that virtual world had on their offline lives. Their ages ranged from 11 to 64 with a mean of 26. All age groups played an average of 22 hours per week. Most (85%) were male,

---

5 http://www.mmodata.net, accessed 26th November 2011
however this varied with age, with female users tending to be older. A third were married and a quarter had children. Many played with a romantic partner or family member. Yee concluded:

The overall demographic composition of MMORPG users is quite diverse, and does not consist primarily of adolescents. In fact, it also includes college students, early adult professionals, middle-aged homemakers, as well as retirees” (2006, p. 327).

Researchers have conducted ethnographies of MMOs (Taylor, 2006; Nardi and Harris, 2006; Ducheneaut et al., 2006; Bainbridge, 2010; Nardi, 2010), as well as “social” virtual worlds (Taylor, 1999; Boellstorff, 2008). Others have examined how MMOG users behave and communicate with each other (e.g. Seay et al., 2004; Moore et al., 2007), and how they use, customize and relate to their avatars (e.g. Ducheneaut et al., 2004; Ducheneaut et al., 2009, Peachey and Childs, 2011).

A key activity in World of Warcraft is the raid, in which a group of players, usually members of the same guild, cooperatively attack an AI monster (‘boss’) in order to receive rewards (‘loot’) that cannot be obtained by a player acting alone. Golub (2010) vividly described the complex and coordinated sequences of actions that must be learned and then enacted in order to defeat a high-level boss:

Technically challenging, phenomenologically intense, emotionally compelling, and deeply connected with self-esteem and group membership, raiding involves serious investments of time and effort, and as a result, successful downing of major bosses is a collective accomplishment that creates social solidarity and can even serve as an important moment in the biographies of individual players. (p. 26)

To achieve this degree of coordination amidst the confusion of battle (and with team-mates not co-present) requires precise coordination, and many high-achieving guilds have a military-style organization, with a guild leader and a group of officers controlling on-going guild affairs such as membership and discipline, while raid leaders, junior raid leaders, and class leaders control raids (Golub, 2010).

Identity is important and complex in MMOGs and can involve avatar customization and user-to-user communication. Researchers have explored the nature of the player-avatar relationship. Ducheneaut et al. (2004) found that users typically maintain several separate characters (“alts”) within a given MMOG, and maintain identities across multiple games, leading to a fractured online identity and problems with maintaining reputation. Ducheneaut et al. (2009) discovered that many users saw their avatar as an idealized version of themselves.

Linderoth (2005) studied children playing platform games, supporting the conclusion of Salen and Zimmerman (2003) that avatars can have several functions:
1. A fictive character that you can pretend to be, a role

2. A piece of equipment, a tool which extends the player’s agency in the game activity

3. A part of the players setting, props which can be used as a part of the players’ presentation of self (Linderoth, 2005).

Klevjer (2006) analyzed the role of avatars in single-player games (where there is no need for communication with other players) and concluded that avatars allow users to vicariously experience a simulated world and act upon it. He pointed out that avatars need not be human-shaped and needn’t even be virtual.

This approach connects computer game avatars to a broader category of avatars, from radio-controlled model planes to Lego men and paper dolls. In computer games, the vicarious body can take different forms; a character, a racing car, a rolling ball, a camera, a gun. (p. 9)

Klevjer argued that a primary function of the avatar is to be a ‘navigable camera’ through which the user views the simulation.

MMOGs are undoubtedly gendered spaces. Yee (2006, cited above) found that 85% of players were male. (Taylor and Witkowski found a similar proportion at the LAN event they observed.) Yet in this space many male players choose to represent themselves as female avatars. Yee (2005) in a survey of WoW players found that males were 8 times as likely to “gender-bend” than females: 23% of male players used female avatars while only 3% of female players used male avatars. Since 84% of players were male, this meant that 1 in 2 female avatars were piloted by a male while only 1 in 100 male avatars were piloted by a female.

Speculation about the reasons for this striking imbalance in gender-play is rife among MMOG players: a Google search finds many discussions. Yee suggested that male players preferred the aesthetic appearance of female avatars, and received more help if other males thought they were female. Blinka (2008) in a survey of WoW and Everquest players found similar proportions: 96% of women played a female character, while only 77% of men played a male character. Blinka also measured players’ identification with their avatars, and interestingly found no differences in identification between users with same-gender and opposite-gender avatars. Blinka concluded that these findings:

...do not show that this style of playing may automatically imply experimenting with the player's identity: the avatar of the opposite sex tends to be used as a mere game tool without any influential reflection. (n.p.)

MacCallum-Stewart (2008) asked “Why do men like playing women so much, and how do they understand this role?” She discussed the reasons that players frequently offer, such as aesthetics (“I have to look at my avatar all day”), or which commentators offer, such as sexual
objectification or exploration of sexual orientation. On interviewing WoW players however she found that the reasons were usually mundane and instrumental: choosing one’s avatar’s gender, like choosing its race and class, is viewed by MMOG players as just another software option to tinker with. While some role-play through their avatars, and some feel they create characters as an author does, most see avatars as tools or game-pieces and their selection as no more important than choosing a colour in chess. Defensive gender-players offer reasons that have been normalized through repetition:

Unease does remain. Players affirm that their behaviour is not deviant by continually having the same conversations about why they choose an alternative gender … men like the female avatars, women want to avoid undue, gendered attention and be regarded as gaming equals. These reaffirmations serve not only to normalise cross-gendering, but also to render it safe, heterosexual and emancipating rather than associated with homosexual desire or transgendered desires. (MacCallum-Stewart, 2008; online)

Taylor (2006) examined the lives of MMOG users as they slipped in and out of complex social networks that crossed online and offline space. She questioned the belief that playing computer games is an isolating and alienating activity indulged in by solitary teenagers, arguing instead that MMOGs are fundamentally social spaces. A number of authors have applied Ray Oldenburg’s (1989) concept of a “third place” to large online games (e.g. Wadley et al., 2003; Steinkuehler and Williams, 2006), to argue that fundamentally these are places for socializing. On the other hand, Ducheneaut et al. (2006) found that much MMOG use was solitary and that it was the ambient social presence of other people, that is the presence of an audience rather than direct interaction, that attracted people to use MMOGs.

Moore et al. (2007) used virtual ethnography, ethnomethodology and conversation analysis to evaluate a range of massively multiplayer virtual worlds for their potential to support collaboration. They focussed specifically on the mechanics of avatar interaction, asking: “to what level of granularity do player activities need to be represented publicly in order to adequately provide for their accountability and tight coordination?” They noted that advancing graphics power had allowed MMOG developers to offer increasingly realistic-looking avatars and environments, yet the level of interactional realism was still crude. The unfolding of turns-at-talk was not sufficiently public and accountable for unproblematic interaction. They recognized however that the environments themselves need not be completely realistic, since to be entertaining there must be an element of fantasy in game scenarios and in what characters are able to do.
2.2.5 ‘Social’ virtual worlds

Virtual worlds are usually categorized into two types: those that are games (the MMOGs) and those that are not (the social or open-ended worlds). As discussed in section 1.1, the term ‘social world’ has been problematized by Lehdonvirta (2010), and is not to be confused with Strauss’ homonymous term. However I have used ‘social world’ because it is popular with the user community. The framework of Schultze and Rennecker (2007) classifies virtual worlds along two dimensions: fantasy-realism and progression-emergence, with the latter dimension discriminating games from non-games.

There are diverse activities that comprise MMOG use (Ducheneaut et al., 2006) but MMOG users are still constrained by the need to collaborate and compete within a formal game framework. This constraint does not apply to social VWs, whose users socialize, role-play, communicate, create content, and engage in a wide range of user-initiated activities (which may include user-devised games). User-created content can be traded, leading to large economies around some virtual worlds (Castranova, 2005).

One of the first social worlds was Active Worlds, released in 1995. Many educators used Active Worlds to support virtual classrooms. Huxor (2001) described his creation and use of a long-term virtual office in Active Worlds. Another early example was the French system Le Deuxième Monde (‘Second World’), released in 1997, which was a simulation of Paris within which users could move and interact.

The social world There ran from 1998 to 2010. This was explicitly a ‘hang-out space’ rather than a competitive game and featured user-modified avatars, the ability to manipulate objects, and sophisticated location-based text chat. The emphasis was on social interaction around activities such as simulated driving, water-skiing and skydiving. Brown and Bell (2004) studied the text-based communication tools in There. Like Moore et al. (2007) they highlighted the problem of aligning avatar action in the virtual world with user action in the physical world. For example, while a user is, say, checking email in another window, their avatar does not reflect this and simply stands lifeless.

The most popular social VW is Second Life (Ondrejka, 2004). This system is discussed in more detail below.
2.2.6 Using virtual worlds to support work

A principal concern of research into computer-supported cooperative work is how to provide each member of a virtual team with sufficient awareness of their collaborators’ actions and context, in order that participants can articulate and coordinate their shared work (Schmidt and Simone, 1996). McGrath and Prinz (2001) examined how co-workers use physical space to articulate work and asked whether virtual space could support a similar process. From the outset it was envisaged that CVEs would support collaborative work. Virtual worlds provide collaborators with a common “place” and would seem to offer a platform that is uniquely able to convey mutual awareness.

Following the recent success of VWs as recreational tools, vendors and users have attempted to adapt them for instrumental uses such as meetings, education and e-commerce. The recent ‘Enterprise Guide to Virtual Worlds’ (Kaye, 2011) lists over fifty such systems. A prominent example is Project Wonderland, developed at Sun Microsystems (Kaplan and Yankelovich, 2011), which offers spatial voice, shared desktop applications, support for voice telephony, and avatars that appear business-like and are tagged with their users’ real names.

Sun used Wonderland to build the MPK20 virtual environment for its own employees, a high proportion of whom are not physically present on the company’s main campus at a given moment. Gardner et al. (2008) used Wonderland to build a mixed-reality teaching environment. Yankelovich et al. (2009) list applications to support virtual whiteboards, remote server administration, and real-time factory process control. Madathil and Greenstein (2011) used it to conduct usability trials of websites with remote participants.

One of the purposes to which virtual worlds have been enthusiastically put is education and training. For this, VWs offer the compelling combination of a simulation framework with online access:

- From learning a foreign language to medical surgery to how to negotiate with a client or implement strategy, virtual worlds foster participatory, collaborative learning in a safe and engaging environment. They allow users to take on different roles in simulated surroundings and to learn from each other through conducting projects or tasks that might be impossible for the single learner or, for that matter, impossible to conduct in the physical world. (Wasko et al, 2011, p. 650).

Many universities have experimented with the use of VWs, in particular Second Life, as teaching spaces: see for example the study of teacher expectations by Saeed et al. (2008), and the survey of use in universities by Gregory et al. (2011).
2.2.7 Presence and immersion in virtual worlds

Presence was defined by Lombard and Ditton (1997) as “an illusion that a mediated experience is not mediated”. They listed technologies that can induce presence:

Virtual reality, simulation rides, home theatre, 3-D IMAX films, state-of-the-art video conferencing, computers that "talk" ... the video telephone, high definition television (HDTV), home and arcade video games, the World Wide Web (WWW), distance learning and telemedicine (Lombard and Ditton, 1997, p. 1).

In the context of virtual environments, which offer both a simulated space and user-to-user communication, researchers differentiate ‘presence’ – that feeling that one is in another (virtual) place - from ‘co-presence’, the feeling that one is in that place with one’s (online) collaborators. Schroeder (2010) has called co-presence “being there together”.

Scheumie et al (2001) surveyed hypothesized causes of presence in virtual worlds including the size, resolution and type of the system’s display, embodiment in the form of an avatar, graphical and social realism, and interactivity. They argued that a CVE induces a sense of presence because it “immerses the user’s senses. This sets it apart from other technologies such as television or books.”

Witmer and Singer (1998) argued that presence is related to attention paid, while Biocca (1997) pointed out that while a person is using a virtual space their attention/presence will shift between their physical and virtual environments. Scheumie et al. (2001) acknowledged that while subjective presence might be a worthwhile goal of recreational systems, evidence was lacking for the usefulness of presence in terms either of task performance or emotional influence.

Bailensen et al. (2006) argued that the appearance and behaviour of one’s collaborators’ avatars can induce co-presence. In an experiment involving one-to-one mediated conversations, these authors manipulated separately the appearance and behaviour fidelities of on-screen representations. By using motion-tracking, they could make the behaviour but not the appearance of the avatar match its owner. The experiment compared voice-only, avatar and videoconference conditions: co-presence was lowest for the avatar. Paradoxically self-disclosure was highest when co-presence was lowest: the authors concluded that: “people emote more freely when their avatar does not express those emotions.”

Recent research on presence in CVEs has addressed its application to psychotherapy (e.g. Juan and Perez, 2009), exercise motivation (Multon et al., 2011), education (Allmendinger, 2010), and the influence of violent games on user aggression (Persky and Blascovich, 2008).
The ability of virtual spaces to induce presence is important to the study of mediated communication within virtual worlds, because different modalities - quite apart from their other properties - might differently influence presence. For example Schroeder (2010) felt that text-based communication “is not VR because it does not enhance – but rather detracts from - the sense of presence and co-presence”.

2.2.8 Types of use and users

Many types of people use VWs for many different reasons (Yee, 2006), presenting a challenge for researchers seeking patterns. A taxonomy commonly cited (e.g. Boellstorff, 2008) broadly distinguishes two types of user:

- The ‘immersionist’, for whom the virtual world presents an opportunity to anonymously play a fictitious character, and
- The ‘augmentationist’, who uses a virtual world as a medium to project their offline identity to collaborators whose identities are also known.

Immersionism is allied to identity exploration of the kind studied in other Internet technologies (e.g. Turkle, 1995) and is more common in recreational contexts. Augmentationism is more likely to be found in workplace, education and commercial contexts.

This taxonony is popular in the Second Life community; for example prominent blogger Gwyneth Llewelyn6 (2008) discussed the terms, claiming that there was a long-term trend towards augmentationist use. However she felt that most people operated somewhere between these two extremes.

Neustaedter and Fedorovskaya (2009a) arrayed Second Life users along a “continuum of immersion” from 'lifer' to 'casual'. ‘Lifers’ were significantly and permanently immersed, and came to the VW to fulfill social needs and obligations. ‘Casuals’ were not as immersed, yet still could form either strong or weak relationships with other users. Lifers, though not casuals, typically had a home, job or romantic partner in Second Life. These user types did not differ in terms of age, duration or frequency of play, and the authors found that some 'casuals' logged in as often as some 'lifers'.

6 This is her Second Life pseudonym.
For MMOGs and MUDs too, frameworks have been proposed for understanding the motivations and activities of users. Bartle (1996) categorized users as achievers, socializers, explorers or killers, expanding this later (2003) to eight categories: friend, griever, hacker, networker, opportunist, planner, politician, and scientist.

Yee (2006) administered a survey asking users of several MMOGs why they played, then performed factor-analysis on the responses to isolate ten components within three the major groupings of Achievement, Social and Immersion. ‘Social’ users enjoyed the teamwork and relationships they found in the VW, while ‘Immersion’ users enjoyed role-play, customizing their avatar and exploring the virtual setting.

Wang et al. (2011) matched player log data with survey data to explore relationships between personality/motivation types, game success, and usage patterns in the MMOG Everquest 2. Through factor analysis this group confirmed Yee’s player types of Achievement, Social and Immersion. Comparing player types with server log data, they found that Achievement and Social players had higher game expertise than Immersion players. Experts also spent more time in-world on completing tasks.

2.2.9 Spatial interaction

Benford et al. (1994) noted that our awareness of other people in space allows us to predict and react to their activities, such as when managing conversations. For example this awareness allows us to see who is interacting with whom, who is available or unavailable for interaction, and whether someone is moving towards us with the intention of interaction. These authors posed questions for CVE design such as: How can a user’s activities and focus of attention be represented to others? How should gesture and facial expression, especially involuntary expressions, be represented? How can an away-from-keyboard user’s lack of presence/attention be represented? How can the history of a user’s activity be represented, since avatars are located “here and now”? Should a user be able to manipulate or block the view they have of another? How can users be represented across multiple media, including audio? Can a user be in several places at once? How closely should an avatar (be forced to) reflect its owner’s true appearance and actions?

These authors proposed a model in which objects (including users) have a location and can interact with each other through a medium. An object’s ‘aura’ is the subspace that defines the object’s presence in the virtual space: that is, the area within which other objects can interact.
with it. Auras can vary with medium: for example the spaces in which an object is visible, audible and touchable might be different. The awareness that one object has of another is defined not only by the latter’s aura but by the former’s ‘focus’ - the part of space in which it directs its attention. An object also has a ‘nimbus’, the region of space in which it tries to project itself in order to attract attention and be noticed. Aura, nimbus and focus can be blocked by boundaries such as walls, and manipulated by adjusting one’s view, zooming in etc. These properties can be used to, say, control the volume levels of transmitted speech, or to define a threshold beyond which text messages are not sent.

Proxemics is the study of how people position and orient their bodies in space during social interaction (Hall, 1959). The existence of avatars allows VW users to simulate physical-world proxemics. Bowers et al. (1996) and Becker and Mark (2002) studied small-group collaboration, noting that users obeyed offline rules of conversation turn-taking and proxemics. Yee et al. (2007) found that Second Life avatars obey proxemic rules even to the point of reproducing gender differences in interpersonal distance.

CVEs might be expected to support collaboration-at-a-distance in problem domains that involve collaboration around objects in space. Domains that have been proposed include architecture, mechanical design, tele-medicine, training, and office work (Van Nederveen, 2007; Rosenman et al., 2006; Satava, R., 1995; Morris et al., 2004; Robinson et al., 2001). A key feature of collaboration in space is that people use their environment to structure their work, referring to objects and locations via gestures and deictic references such as “this thing” and “that place”. Yet collaboration in virtual environments has proven to be problematic.

A central issue has been how a user can deduce another user’s point-of-view in order to reference objects deictically. Hindmarsh et al. (1998) found that groups using desktop CVEs struggled to achieve common reference to objects, even when their avatars were able to point, primarily through failing to see both the pointing arm and the referent due to the narrow field of view. Performing virtual gestures also forced users to spend too much time ‘driving the avatar’. To make up for the shortfall in awareness, people used more verbal communication than would be expected of collaborators in physical environments.

In another study it was found that collaborators ignored the spatiality of the VW, degrading the system to a simple chat tool. “Each user would find no reason to move their avatar, and would simply engage in text-chat from wherever they found themselves. The machine was rendering a 3D space for no real purpose” (Huxor 2001, p 285). Similarly in Evard et al.’s (2001) study of workplace use of a MUD-based social world, users worked around the requirement to be in
the same virtual location in order to converse, by creating multiple characters in order to be simultaneously present in many virtual rooms.

“People can either be actively taking part in multiple conversations in different rooms, or simply having their character(s) listen in’ – that is, having a character in a room means one can record relevant conversations in the form of logs for asynchronous review” (Evard et al., 2001; p 278).

Collaboration and conversation in virtual space remains problematic and is examined in studies two and three.

2.2.10 Second Life

One of the most prominent ‘social’ worlds is Second Life, developed and marketed by San Francisco company Linden Lab (Ondrejka, 2004). As a non-game virtual world, use is not structured by competition or combat. Second Life (SL) was the platform for two of my studies.

The contents of SL are built by its users. User constructions are collections of simple shapes called ‘prims’. Building is a popular activity, with dwelling construction one of the most popular projects. User accounts are either paid or free: of a total of about 1.5 million users, about 80,000 pay and can own virtual land and build on it. The right to build permanent structures is the main advantage conferred by a paid account, so this is a reasonable estimate of how many users actually create content. Creations can be sold from user to user, paid for in the ‘linden’ currency which can be exchanged for US dollars. The exchange of (in this case, user-created) objects for money creates a large economy around this virtual world, just as it does in many MMOGs (Castronova, 2005).

Second Life provides no specific support for collaborative building – the focus of my third study - beyond the ability to visualize one’s collaborators in the shared workspace (cf. Seay et al., 2004), and to communicate with them using text or voice. Users can display a map of the local area showing avatar locations and the outlines of buildings; however the map is not sufficiently detailed to assist with object manipulation and tends to be used only for coarse navigation. A number of authors have commented on SL’s potential as a tool for computer-supported cooperative work: Van Nederveen (2007) proposed it be used for collaborative architectural design, while Rosenman et al. (2006) tested a design system in which SL was supplemented with tools such as a 2D sketch-pad.

The SL client software uses a standard PC screen for output, and a keyboard and mouse for input. Since 2007 a spatial voice channel has been included for user communication: this is the
focus of my fourth study. Users can choose between first-person (through the avatar’s eyes) and third-person (from behind the avatar) views.

Unusually, SL allows a user to move their camera (the point from where the user’s view is calculated) independently of the position and orientation of their avatar, by anchoring it to an object in the local scene. This technique, called a “virtual camera” by Hindmarsh et al. (1998), allows a user to gain multiple perspectives of an object more quickly than is possible by walking their avatar around the object. Thus it is commonly used during object-focused activity such as building and looking at other users’ creations. Unlike most VWs, which support limited camera movement near the avatar, the SL camera can be moved over a wide area, oriented in any direction including up and down, zoomed a long way in and out, and unlike avatars, moved through walls. While users’ avatars are publicly visible, their camera positions are not (Irani et al., 2008). The ability to decouple one’s camera from one’s avatar is similar to techniques suggested by Hindmarsh (1998) and Bailensen et al. (2006).

Second Life’s virtual camera means that there is no reliable relationship between what an SL user can see and what their avatar appears to be looking at. While the feature is useful when editing, prior research (e.g. Hindmarsh et al., 1998) suggests that when users cannot deduce each others’ vistas, their ability to collaborate is lessened.

Second Life users must use a pseudonym: this is displayed above their avatar and is searchable in a database (illustrated in figure 1, chapter 1). There is a social prohibition in SL against linking a user’s offline and SL names, and during in-world conversation users usually refer only to SL names (see conversation fragments in study 4 for examples). Each SL user can maintain a friends list, similar to those used in social network sites, to keep track of other users with whom they regularly interact. SL also offers a group feature with membership control. Groups are a popular feature in SL and are used to organize discussions and other in-world activities.

Two types of text communication are offered in SL. Messages (called ‘Instant Message’ or ‘IM’) can be sent to users on one’s friends list. Users can also simply type text into a chat box: this is sent to everyone who has an avatar nearby.

Second Life voice can be used in several modes. The default is a spatial channel which streams audio between users depending on their avatars’ proximity. Users can also set up one-on-one ‘telephone calls’ or group ‘conference calls’, and it is possible to connect with non-SL-users via regular telephone services. Linden Lab stated in 2009 that voice was used by 50% of its users and made “education, business meetings, simulations, training and casual conversations

Second Life offers pre-programmed gestures which can be triggered by keystrokes or by selection from a menu, and allows users to program their own gestures as well. However avatar gesture is of limited usefulness as it requires conscious effort by the sending user. Use of gestures in SL is typically restricted to occasionally waving ‘hello’. Some research has attempted to address this problem, for example by using cameras to recognize a user’s unconscious body movements and translate these to the avatar (e.g. the Hands Free 3d project at http://www.handsfree3d.com/).

Second Life was promoted as a forum in which corporations could interact with customers through advertising, events and the sale of virtual goods. Many companies built a virtual presence in SL; however this experiment is generally considered to have failed. Despite Gartner’s earlier prediction that virtual worlds would become important to business (Gartner, 2007), it has been suggested more recently that VWs have entered a “trough of disillusionment” with regards to business use (Salomon, 2009). The failure to connect with customers may be partly due to the large size of the virtual space relative to the number of users, so that: “the chances of would-be customers seeing a company’s Second Life branding, much less interacting with it, are often slim” (King, 2008). On the other hand, Wasko et al. (2011) recently argued that virtual worlds have emerged from the trough of disillusionment onto the ‘slope of enlightenment’ where real rather than speculated benefits accrue.

IBM made significant investments in Second Life and in its own virtual world technology, delivering products such as Metaverse and BlueGrass, designating employees as ‘metaverse evangelists’, and issuing a guide to employee conduct in virtual worlds. An interview with one Metaverse leader revealed IBM’s interest in the viability of business-to-customer e-commerce. For example clothing and shoe companies were expected to sell virtual items for use by avatars, which would then translate to sales of the physical items.

A striking aspect of Second Life relative to other virtual worlds and especially to MMOGs is the range of uses to which it has been put: these include online conferences (e.g. Erickson et al., 2011), teaching (e.g. Gregory et al., 2011), promoting health (e.g. Cowdery et al., 2011) and psychotherapy (Gorini et al., 2008).

---

7 accessed at http://www.3d-test.com/interviews/ibm_1.htm
Harris et al. (2009) individually tracked the behaviour of 80 university students as they used *Second Life* over a period of 6 weeks. The activities pursued, in order of popularity, were parties, educational activities, cultural events (music, museums, poetry), gambling, sporting events, sexual activities, religious ceremonies, and community outreach. Over time these novice users spent more time in populated areas and expanded their number of friends and groups. However they chatted less, possibly due to the ‘alone together’ syndrome proposed by Ducheneaut et al. (2006), and possibly because they required less help from each other in using SL. (This study measured only text, not voice, use.)

Green-Hamann et al. (2011) studied social support groups relating to alcoholism and cancer care who met regularly in *Second Life*. They found that the combination of avatars and nearly-synchronous text communication provided an appropriate balance of anonymity and social cues, placing this form of interaction somewhere between face-to-face meetings and web forums. (These groups used text rather than voice.) Avatars and text enabled “intimate, hyperpersonal relationships with people who care about them and who help them cope with their stressful life circumstances”.

### 2.2.11 Analyzing interaction in virtual worlds

A number of researchers have employed conversation or interaction analysis to study user-to-user interactions in virtual worlds.

Conversation analysis (CA) emerged from attempts by Garfinkel, Sacks, Schegloff and Jefferson to devise a rigorous, objective and scientific method for studying social interaction, and was enabled in particular by the invention of audio and video recording devices (ten Have, 1999). Psathas (1995, p. 10) felt that CA’s emphasis on rigour was in response to the popularity of qualitative, observational approaches such as Goffman’s, as well as to quantitative methods that placed utterances into pre-formed categories and ignored the details of interaction.

CA was influenced by ethnomethodology and shares with it the core belief that human interaction has an order which is produced by the interactants, and can be discovered through analysis: “Social actions in the world of everyday life are practical actions and are to be examined as ongoing practical accomplishments” (Psathas, 1995, p. 3). Thus the phenomena that CA seeks to study include not just conversation but interaction, so that the term ‘interaction analysis’ is preferred by some.
Fundamental to conversation is the taking of turns: that is, members of a conversation group do not usually speak over each other, but wait while others are speaking. The mechanisms by which turns are requested, allocated and sequenced are key phenomena in CA. Conversation structure is ad-hoc, and coordination problems are repaired by participants as they occur.

Within CSCW research, conversation analysis has been used to examine interactions among co-workers, such as in Heath and Luff’s (1991) study of collaboration in railway control rooms. Because interaction over the Internet “overwhelmingly takes place by means of discourse” (Herring, 2004, p. 339), it has been amenable to a form of CA which Herring (2004) called “computer-mediated discourse analysis”. Again, the aim has been to “employ micro-level methods to shed light on macro-level phenomena” (p. 338).

Bowers et al. (1996) used CA to study a project meeting held in one of the first networked virtual worlds, MASSIVE. They uncovered significant problems with turn-taking, due to the lack of bodily cues such as glancing which normally occur in f2f conversation, and despite the embodiment of users as avatars. Turns-at-talk often overlapped, or contained long gaps. The authors argued that whereas silence in f2f is often socially meaningful, in a VW it is more likely to be caused by technical difficulties.

Hindmarsh et al. (2000) used CA to study object-focused collaboration in the same system, finding that avatars, even when they offered the ability to point, did not much help users understand their collaborators’ deictic references. Conversation analysis showed that a great deal of talk was devoted to repairing interaction.

Brown and Bell (2004) studied the virtual world There, using two methods. They conducted virtual ethnography “with a significant amount of time spent online in the research setting, observing, participating and taking field notes”, and video-recorded social interactions in order to conduct CA in the style of Heath and Luff. They found that users engaged in significant interaction around objects. Overlapping conversation turns occurred due to the way the system displayed unfolding text messages: users exploited this to hasten conversations and tell jokes.

Schroeder et al. (2006) compared CA with a quantitative method in the context of research into CVE usability. The quantitative method involved categorizing interactions in a recorded session into categories such as Communicate, Manipulate and Navigate. This allowed the researchers to compare the frequency of interaction types across different users and tasks. They compared this approach with a variant of CA which combined transcripts of conversation with maps of the path navigated by users through the virtual world. The researchers felt that for CVE research, CA had the advantage of showing in detail where usability problems arose. However it shared with ethnographic description the disadvantage of doubtful generalizability.
They concluded that a combination of the two methods might work best. (This is the approach I took in study 3.)

Moore et al. (2007) analysed interactions between MMOG players to discover how well avatar systems approximated face-to-face interaction. They video-recorded player-to-player interactions and selected salient phenomena for transcription. They were interested in questions such as: How is turn-taking in voice communication organized? How are avatar orientation and gesturing used? How can users determine where another user is looking? They focused on whether avatars made user actions ‘accountable’ (in Garfinkle’s sense) sufficiently to support coordination. Some problems were caused by the text medium: turns and pauses were longer because typing takes longer and people could not see their interlocutors’ turns unfolding, as messages were only sent after they were completely typed. They found that MMOG avatars do not “give off” (in Goffman’s sense) as much information about their owner’s actions as real bodies do: avatars typically stood idle while their owners performed actions such as checking inventories and selecting menu options. They also found that avatar eye-gaze did not correspond with what users could see in most MMOGs, and that gestures were difficult to perform (requiring menu use) and could not be ‘held’ for coordination purposes but instead ran for a fixed period of time.

2.2.12 Conclusion

In section 2.2 I have reviewed research into virtual worlds, highlighting important issues such as presence, collaboration, and taxonomies of use and user. In particular I described the differences between the two major types of virtual world - massively multiplayer online games and social worlds - and discussed methods that have been used to study them.

Users of virtual worlds are usually not physically co-located, and therefore must use a medium in order to communicate linguistically with each other. We saw above that VWs typically support this by providing text and/or voice subsystems. I will show in section 2.4 that few studies have explicitly compared the use of modalities in virtual worlds. However a great deal of research has performed such comparisons outside the VW context: this work is reviewed now, in section 2.3.
2.3 Mediated communication

Fairclough (2003, p. 77) distinguished four kinds of linguistic communication:

1. Two-way non-mediated (face-to-face conversation)
2. Two-way mediated (e.g. telephone, email, video-conferencing)
3. One-way non-mediated (e.g. lectures)
4. One-way mediated (e.g. print, radio, television, film)

The scope of my research is limited to the second of these types: communication via two-way media. I am concerned with the effects that properties of different media have upon communication. However before examining media and their influence it is necessary to begin with an understanding of the nature of human communication. For this I have adopted the framework of Clark (1996).

2.3.1 Use of language

Clark (following philosophers such as Austin and Searle) emphasized that language is used by people to do things. This is a focus on the use of, rather than production of, language. In particular he proposed language to be a sub-category of *joint action*, which is performed by (often temporary) groups which Clark called “ensembles” (1996, p. 3). Clark analysed and categorized joint actions, using a form of interaction analysis. He argued that as a joint action, language comprises both individual and social processes and is not reducible to either of these.

Clark classified language settings according to medium and scene. Medium is the most fundamental categorization and comprises spoken and written settings. The range of scenes includes:

- Personal (regular conversation)
- Non-personal (monologues, lectures, newspaper articles)
- Institutional (structured, as in legal proceedings and business correspondence)
- Fictional (e.g. plays and novels)
Mediated (note that Clark means speaking through an intermediary such as a secretary, not through a medium such as a telephone)

Private (speech and notes to self)

Clark’s dimensions of medium and scene are orthogonal, so each of these scenes can involve either spoken or written language.

Clark considered face-to-face conversation among small groups to be primal, or in Clark’s words “the cradle of language use” (1996, p. 99), that is, this form of conversation appears before others in both developmental and historical senses. Understanding face-to-face conversation is therefore fundamental to understanding mediated communication, and analysis of the mediated situation is often based on a comparison with face-to-face. Face-to-face conversation is characterized by co-presence, mutual visibility and audibility, instantaneity (lack of delay in perception of actions), evanescence and recordlessness (utterances fade quickly and are not preserved), simultaneity (face-to-face is a two-way channel), extemporaneity and self-determination (participants control their own actions in real time without a plan), and self-expression (the participants act as themselves, and are not, for example, acting a role).

Clark held that the basis of all communication, indeed all joint action, was common ground: the ensemble’s collection of shared knowledge, beliefs and understandings, including their common language and their (presumably shared) perception of activities and objects in their immediate context. In fact Clark viewed conversation as a process whereby a group accumulates common ground (p. 39).

Clark provided a framework for understanding the participants in a conversation which is useful in this thesis. Imagine 5 people labelled A to E. A, B and C are talking and consider themselves to be “ratified participants” in the conversation. While A (speaker) is addressing B (addressee), C is a “side participant” who is expected to be following A’s utterances even though they are not directed specifically to C. D, a “bystander” is not part of the conversation group but is located near enough to overhear them, and for the group to be aware that D can hear them. E, an “eavesdropper”, can also hear the conversation; however A, B and C are not aware of this. (Research into mediated communication reviewed below has shown that the possible presence of eavesdroppers influences choice and use of media.)

During conversation people often have cause to describe hypothetical or fictional situations, such as when telling jokes, acting, or writing stories. To analyse these situations Clark used the metaphor of layers, which are somewhat like theatre stages. People use language in the here-
and-now (layer 1) to support a story involving hypothetical characters and actions (in layer 2), and recursively more layers can be used.

Clark analysed the dimensions of joint actions (the superset of language use) as:

- scripted vs unscripted
- formal vs informal
- verbal vs non-verbal
- cooperative vs competitive
- egalitarian vs autocratic (1996, p. 31).

(The joint activities of virtual world users can be analyzed along these dimensions also, and occupy a variety of points along most of them, although their actions are mediated and occur within a virtual space.)

Clark felt that “how people coordinate is one of the fundamental issues of language use” (p. 36), leading him to use a form of conversation analysis in his work. He argued that ultimately people coordinate via signals and that these are not always linguistic. In face-to-face conversation, speech is typically interwoven with a rich array of gestures. These are integral to speech, yet are frequently ignored in language analysis: “non-linguistic signals are subtle, highly systematic, and not at all ad hoc” (p. 156). There are many kinds of non-linguistic signals including iconic, indexical, and demonstrative gestures. Human signals can utilize the voice, hands, arms, face, eyes and body (p. 188).

### 2.3.2 Communication technologies

Communication technologies free people from the need to be physically and temporally co-located in order to communicate with each other. This is an ancient concern. Nineteenth and twentieth century innovations such as the telegraph (Standage, 1998) and telephone (Fischer, 1994) enabled conversation at a distance. Conversation via video link become possible in the mid 20th century, though it was not widely used until recently (Roussel and Gueddana, 2007). Computer-mediated communication (CMC) became available in the late twentieth century and since the popularisation of the Internet enjoys widespread use.
Each communication medium has properties which influence both the decisions of people to use that medium, and the quality and nature of communication that people subsequently accomplish with the medium.

Virtual worlds, of course, are a kind of communication technology. But I argue in this thesis that VWs are unique in that they simulate a space, locate users within the space, then provide them with media with which to communicate within the space. I call these media-within-media ‘virtual media’, because they are typically designed to resemble the familiar text- and voice-based media used outside the VW context.

2.3.3 Social presence

A key publication on the influence of media properties on communication was Short et al.’s (1976) *The Social Psychology of Telecommunications*. This work preceded the widespread use of computer networks for communication, but covered early experiments with video-conferencing as well as telephone use. The authors reviewed research comparing audio and video communication to face-to-face conversation with respect to criteria such as efficiency, persuasiveness, conflict resolution and group cohesion. The authors were concerned primarily with how well, in a business context, mediated communication would fare relative to face-to-face meetings.

These authors proposed the *social presence* theory of mediated communication. Social presence was defined as information about the interlocutors, transmitted through the medium alongside the linguistic message that the interlocutors intended. In other words, while media are designed to transmit linguistic messages, some whether by intention or a side-effect of their design also transmit personal information about the person sending the message. Short et al. arranged media according to how much social presence each conveyed – this rank corresponding to the medium’s *richness*.

The authors held that face-to-face conversation conveyed the maximum amount of social presence. Interlocutors could hear, see, and even smell or touch each other, and thus receive the maximum amount of information about their conversation partner that could be conveyed. They might, for example, deduce the age, gender or nationality of an interlocutor, their mood, or their attitude to the topic being discussed.
According to social presence theory, the medium that conveyed most presence was video-conferencing, as it allowed interlocutors to see and hear each other and, for example, perceive facial expressions along with speech.

Audio-conferencing and telephone conversations were next in the social presence hierarchy: these allowed people to hear the quality of each other’s voices, which provided cues - less than did video - about the person speaking.

Finally, written documents conveyed only words and thus the least social presence.

The motive for this research was primarily economic. The low social presence conveyed by mediated communication was seen as a cost to be weighed against that of transporting attendees to meetings. That is to say, the absence of social presence that a given medium might project was seen as a disadvantage, not believed to confer any benefit other than cheapness. This view has been critiqued in subsequent research, reviewed below.

The social presence theory was prescriptive in that it suggested that people ought to choose the medium that provided the degree of social presence required for the communication task at hand (and no more – so as to choose the cheapest feasible technology).

### 2.3.4 Media Richness Theory

Daft and Lengel (1986) extended the social presence theory of Short et al. (1976) in several ways:

- the properties of a medium that determine its richness were analyzed,
- the conditions under which richer media should be used were defined, and
- early forms of (text-only) computer-mediated communication were analysed.

The uni-dimensional hierarchy of media was preserved, with face-to-face conversation - the richest - at the top, followed by video conferencing, then telephony, CMC and finally unaddressed written documents at the bottom (Daft et al. 1987).

The focus of Media Richness Theory (MRT) was both on the amount of information that a medium can transmit (using Shannon and Weaver’s 1949 definition of information as “that which reduces uncertainty”) and on how well the medium lent itself to the resolution of equivocality, that is, the efficient discussion of multiple possible interpretations of data. The
face-to-face condition and rich media allowed people to efficiently arrive at consensus. Lean media did not as readily support brisk conversation and a process of resolution.

Daft and Lengel listed four properties of a medium which determine its richness:

- the ability it afforded interlocutors to give instant feedback,
- the delivery of multiple cues (e.g. gestures and tone as well as words and numbers),
- the variety of language symbols that could be transmitted, and
- the ability provided to interlocutors to transmit personal feelings and emotions.

The authors proposed that a medium of appropriate richness should be chosen for each situation. The richest was not always necessarily the best. When, for example, a large quantity of objective facts needed to be transmitted, personal cues could be distracting and a lean medium would be superior (Daft et al. 1987). They felt that a preference for rich media was one reason that managers at that time were not adopting computer-mediated communication such as email to the extent many had expected. They recognized that email combined the instant feedback of telephone conversation with the ability to reach a large, geographically-dispersed audience, though their survey found that: “electronic mail is considered inappropriate for exchanging confidential information, resolving disagreements, getting to know someone or negotiating”.

Media Richness Theory shared some of the limitations of the earlier Social Presence theory:

- Media were analysed with respect to their usefulness only in business contexts. Recreational CMC has proven to be problematic for MRT.
- The theory was partly prescriptive of which media an organization ought to use in particular situations, and partly descriptive of what people actually do.

### 2.3.5 Responses to MRT

Broadly, MRT has prompted two responses. One has been the ongoing development of communication tools with ever-greater richness: this is the project of attempting “to imitate face-to-face” (Hollan and Stornetta, 1992). The other response has been to critique the claims of MRT. Critiques were proposed when it was found that:

- people sometimes choose media by criteria other than richness,
- people sometimes choose lean media when richer alternatives are available, and
users of lean media often achieve rich communication regardless.

These findings have prompted alternative models of media choice. For example, the social-influence (Fulk et al., 1990) and critical-mass (Markus, 1990) frameworks proposed that an individual’s choices of medium was influenced less by the medium’s properties than by the choices of the individual’s collaborators. Other researchers argued for a shift in focus from people’s choice of individual media to their use of a collection of media (Watson-Manheim and Belanger, 2007). It was found that people linked by strong ties communicated with each other via a greater number of media (Haythornthwaite, 2002). The size of the group communicating is an important factor influencing the usefulness of a medium: in a series of experiments comparing voice-over-IP with instant messaging, Lober et al. (2007) found that text scaled better with group size than did voice.

Other researchers have focused on the characteristics of the individuals doing the communicating. Carlson and Zmud (1999) argued that communication richness depends on its users’ prior experience with the medium they are using and with each other. Walther (1992) argued that people in particular relationships or situations are motivated to work around the shortcomings of lean media and create rich communication. Goby (2006) found that different personality types had different preferences for online vs face-to-face interaction, with shy people preferring leaner media.

Rich communication involving a strong sense of one’s interlocutors can be achieved without the use of rich media. Ngwenyama and Lee (1997) found that users of lean media bear the identity of a sender in mind to interpret the validity of messages they receive, and argued that richness is produced by the social context of the people conversing. Nardi et al. (2000) found that workplace users of instant text messaging maintained a sense of connection with each other, achieving in effect a kind of lean media space. Vetere et al. (2009) argued that many messages exchanged for the purpose of maintaining intimate relationships are not informational but phatic: intended to keep the communication channel open.

An influential rejoinder to MRT was presented by Hollan and Stornetta (1992). Critiquing attempts to build video-conferencing systems of increasingly high fidelity, these authors argued that the project to invent media that simulate as near as possible the face-to-face condition was a flawed approach. They felt that mediated communication would never be ‘as good as the real thing’, so technology should instead attempt to be better. They suggested one would know that this had been achieved when people who were physically proximate, and could therefore meet face-to-face, chose to use a technology instead. (They noted that email, a lean medium, was already being used in circumstances like this.)
Hollan and Stornetta argued that, to achieve this end, researchers should find out what is essential and what is inconvenient in face-to-face communication, in order to support and enhance the former while omitting the latter. They listed problems with face-to-face that a future technology might attempt to fix, such as the difficulty of archiving spoken language, lack of clarity, and lack of precision in gestural feedback.

### 2.3.6 CMC and online communities

Developments in computer-mediated communication presented challenges to MRT and prompted new theories. Dennis et al. (2008) developed the Media Synchronicity Theory, which analysed media according to ‘capabilities’ enabled by the physical properties that a medium may possess. These included:

1. transmission velocity,
2. symbol set,
3. parallelism,
4. rehearsability, and
5. reprocessability.

(The first two criteria were equivalent to the ‘instant feedback’ and ‘symbol variety’ criteria of MRT.) For example, while email offers a limited symbol set and transmission velocity, it supports large numbers of users working in parallel, and allows them to rehearse (edit) a message before transmitting, and to store and re-read messages.

Widespread adoption of computer-mediated communication has allowed the formation of online communities: groups of people who use CMC as their primary medium of communication. Fernback and Thomson (1995) defined an online community as a 'social relationship forged in cyberspace through repeated contact within a specified boundary or place (e.g. a conference or chat line) that is symbolically delineated by topic of interest'. Rheingold (2000) felt that online communities could fulfil social needs:

Perhaps cyberspace is one of the informal places where people can rebuild the aspects of community that were lost when the malt shop became a mall. (p. 10)

While there was an early emphasis on the ‘separateness’ of the online world, researchers have since focussed on how people integrate online community within their everyday social networks (Wellman and Haythornthwaite, 2002). Social networking websites have become a
popular genre of CMC which predominantly supports relationships that already exist offline (boyd and Ellison, 2007).

2.3.7 Online identification

Because it does not transmit the user’s voice or appearance, text-based CMC allows the user to construct an identity (Turkle, 1995). Active self-representation through the use of profiles, photos and avatars has afforded complex manipulation of how one is presented to others (Subrahmanyan and Smahel, 2011, p. 59). Internet users can be anonymous or pseudonymous or have a fluid identity.

People may present themselves pseudonymously for a number of reasons, such as the desire for privacy, or in order to deceive others, or because they wish to explore alternative identities. Turkle (1995) felt that identity exploration using the Internet could be a positive, therapeutic experience, and Bartle (2003) reaffirmed this in the context of online role-playing games. Many MMOG users value the use of pseudonyms (Albrechtslund, 2011). But anonymity has ‘positive and negative implications” (Christopherson, 2007).

The ability to verify identity contributes to knowledge sharing within some online communities (Ma and Agarwalm 2007). Law enforcement officials and companies engaging in e-commerce would prefer online identity to be more readily verifiable (Davenport, 2002). The popularity of sites such as Facebook suggests that many people value verifiability over pseudonymity. This remains a contentious issue: see for example the recent debate between 4Chan’s Christopher Poole and Facebook’s Mark Zuckerberg (Guardian, 2011).

In Second Life, pseudonymity has been mandatory or at least encouraged, as users were not permitted under the terms of service to use real names or disclose the names of other users. (This policy has been relaxed recently with the introduction of the “display names” feature.) SL users have extensive control over the appearance of their avatar, and spend significant time customizing it in order to project a desired identity (Neustaedter and Fedorovskaya, 2009b; Ducheneaut et al., 2009). There is a market for the exchange of items such as avatar skin, hair and clothing, which are used to tailor aspects of avatar identity including race, gender and age (Reed and Fitzpatrick, 2008). Some users come together in SL for the purpose of role-playing in a group.
2.3.8 The social context of communication

Technology research has increasingly focused on the context in which use happens (Dourish, 2004). This is partly in reaction to the increasing mobility of technology and the movement of computers from the workplace into social and domestic spaces (Gaver, 2002; Howard et al., 2007).

The context of communication technology use can be especially complex, because co-located others can become unintended parties to a mediated conversation. Here a medium’s leanness can be an advantage. Jones (2002) commented on “the different sets of ‘mutual monitoring possibilities’ which these technologies make available, the different ways in which they allow us to be present to one another and to be aware of other peoples’ presence.” Jones argued that it is not so much the ability to send messages through space that gives text-based CMC its power, but the ability to constrain who receives messages, which allow users to multi-task and multi-converse:

The ‘muting’ of the visual mode, for example, allows users to engage in a wide range of physical activities that are inaccessible to their on-line interlocutors, and the ‘muting’ of the aural mode allows them to carry on on-line conversations which are inaccessible to others who are physically co-present. (Jones, 2002, p. 13)

In other words, on-line conversation does not leak into the offline context, and vice-versa.

People wishing to engage in deception or ‘impression management’ (Carlson et al., 2004), or who are shy (Stritzke et al., 2004), are inclined to choose a lean medium such as text. People value mediated communication for the ability it offers to manage self-presentation (O’Sullivan, 2000), especially in face-threatening situations (Feaster, 2008). Control over self-presentation is not only desired in recreational systems, but is actively exploited in work contexts (Vast, 2007). A major driver of social-network-website use is the ability to present an idealized persona (Dwyer et al., 2007).

Hancock et al. (2009) observed that instant-messaging users employed a form of deception they called ‘butler lies’. The authors found that 10% of IM messages were deceptive, and that of these, about a fifth were butler lies, with ‘exit strategies’ (to end the conversation) the most common. They critiqued the provision of presence technologies (such as video-cameras) designed to broadcast one’s availability for conversation, pointing out that people are not simply 'available or not', but rather are differently available to different people.

Studies such as that of Aarsand (2008) have shown that CMC users exploit opportunities to multitask. Aarsand studied school students in the classroom who alternated between essay-
writing and IM, concluding “there was not one dominant activity frame, but rather the activities involved borderwork, and more specifically frame switches and a strategic use of tags.”

Research into mobile phone practice has discovered that a common motive for using text instead of voice is to avoid eavesdropping in public places (Faulkner and Culwin, 2004). Teenagers seeking privacy (Davie et al., 2004), and employees in meetings (Norman, 2004) choose SMS over voice in order to avoid being overheard. Mobile phone users actively use features of their phones to control who does or does not participate in their social interactions (Madell and Muncer, 2007). Eavesdropping was a common concern in the early days of party-line telephony (Fischer, 1994).

Of course third parties to a conversation do not always want to eavesdrop. McCarthy and Wright (2004) described the experience of being in a train where people speak noisily on mobile phones:

> For those who get irritated, it is not the idea of people talking on their phones in a public space that is annoying. It is the sensory or physical quality of the intrusion. The noise seems to permeate a boundary. The noisier it is or the more grating the ring or the voice, the more violent the intrusion. (McCarthy and Wright, 2004, p. 13)

The social context of (non-mediated) communication, and in particular the work people put into presentation management, was analysed by Goffman (1959), who described the ways in which people actively seek information about each other while interacting. He noted that people deduce clues about identity, socio-economic status and beliefs from each other’s appearance, behaviour and speech. Furthermore, in response to the information-gathering efforts of others, people try to control the impression they are presenting. Goffman differentiated a person’s voluntary and involuntary expression of feelings, calling these “two radically different kinds of sign activity: the expression that he gives, and the expression that he gives off” (1959, p.2). Voluntary expressions include language and formal signs. Involuntary expressions include speech intonation, gestures, and spatial positioning of the body.

In accordance with Krebs and Dawkins (1984), Goffman’s analysis emphasizes that one should not assume cooperation between communicators. This point was emphasized by Bannon and Schmidt in a CSCW context: “We view organizations as a coalition of individuals motivated by individual interests and aspirations and pursuing individual goals” (1989: p368). Goffman’s framework has informed analyses of Internet communication, such as that of Miller (1995).
2.3.9 Computer-Supported Collaborative Work

Research into computer-supported cooperative work is concerned with the design and evaluation of information systems that support groups of people collaborating to achieve a shared goal (Greif, 1988; and see Schmidt and Bannon 1992 for comments on the difficulty of defining ‘group’, ‘shared goal’ and other terms.)

Typical CSCW scenarios involve ad-hoc or permanent groups, some or all of whose members are geographically dispersed. Johansen’s time-space matrix (1988) divides group work into four categories defined by two orthogonal dimensions: whether the collaborators are co-located, and whether the collaboration is synchronous. Collaboration across different locations or times usually requires that a communication technology be used. However Schmidt and Simone (1996) noted that all collaborative work is distributed even if the workers are co-located, because collaborators are “semi-autonomous in terms of the different circumstances they are faced with in their work as well as in terms of their strategies, heuristics, perspectives, goals, motives, etc”.

Computer systems mediate CSCW by transporting messages and status information about the users and their field of work, via modalities such as text, graphics, sound or video (Ellis et al., 1991). A principal concern is how to provide each participant with sufficient awareness of their collaborators’ actions and context, and how to allow participants to articulate and coordinate their shared work (Schmidt and Simone, 1996). Because these functions require effective communication, CSCW and CMC are by necessity related fields.

Key authors in CSCW have insisted that the field be concerned with supporting work and not just group activity in general (Schmidt and Bannon, 1992 and other papers). However other authors have extended the principles of CSCW to inform new fields, such as computer-supported cooperative learning (McConnell, 2000) and computer-supported cooperative play (Ishii et al., 1999; Wadley et al., 2003; Brown and Bell, 2006).

Schmidt and Bannon (1992) described two ways by which a CSCW system can support coordination. One is to automate and control workflow. The other is to offer a “shared information space” which “allow[s] the members of a cooperating ensemble to interact freely, i.e. without being constrained by prescribed procedures or established conversational conventions” (1992: p20). These authors emphasized the need for collaborators to be able to effectively articulate their shared work, quoting Robinson (1991): “In general it can be said that any non-trivial collective activity requires effective communication that allows both ambiguity and clarity.”
Concerns for CSCW include:

**Awareness**

To coordinate joint action, collaborators require awareness of each others’ activities and the status of their field of work. If the collaborators are co-located this information can usually be made available and monitored without much effort on the part of the collaborators (though see Schmidt, 2002, for a discussion of how complex this process can be). Researchers are concerned therefore with how CSCW systems can provide the right quality and quantity of information in order to make mutual awareness feasible for collaborators who are not co-located and possibly not working synchronously.

Dourish and Bellotti (1992) defined awareness as “an understanding of the activities of others, which provides a context for your own activity”. Their emphasis was on passive awareness mechanisms, by which actors “give off” (to use Goffman’s phrase) information about their actions without needing to actively provide it to the system.

Two broad approaches have been made to the facilitation of awareness within CSCW: “media spaces” and “event propagation mechanisms” (Schmidt, 2002). The difference arises from different assumptions about the type of information that facilitates awareness and how it is or should be given and received.

The media-space approach (Gaver et al., 1992; Bly et al., 1992; Heath and Luff, 1991) builds upon the observation that co-located collaborators are able to communicate and monitor each other’s activities in a natural, seemingly unconscious way, as well as “the premise that work is fundamentally social” (Bly et al., 1992). Researchers in this tradition have reasoned that by imitating the face-to-face condition using rich media such as video-conferencing to continuously link remote offices, one could facilitate a natural process of collaboration.

In an early media-space experiment, staff at two Xerox PARC locations used video (as well as email and other technologies) to link individual offices and common areas. Bly et al. (1992) listed the key uses made of this link as awareness, chance encounters, locating colleagues, visual explanation, group discussions, recording and replaying “video notes”, presentations, and social activities. Staff in video-equipped rooms dealt with eavesdropping (and its converse, interruption) simply by switching microphones and cameras on and off. Microphones were typically ‘off’ for much of the time in individual offices, but left on in common areas.
Gaver et al. (1992) evaluated another media space that connected researchers within a multi-story building at EuroPARC and to the PARC lab in California, via video and audio links. The authors argued that collaboration is not merely confined to teams intensely focussed on a common task, but includes relationships with lesser levels of engagement or planning. Therefore they designed their system to provide “general awareness” to enable the serendipitous collaborations that typically occur among co-located workers, as well as fluid movement between degrees of engagement. They focussed on privacy issues and in particular on giving users control over who could see or hear them at a particular time.

The relationship between awareness and media richness was illustrated by Gutwin and Greenberg (1996) who described online workspaces as “impoverished when compared with their physical counterparts”, and spoke of their desire to “enrich” online interaction. However in a special issue of the journal CSCW on awareness, Schmidt (2002) critiqued the approach of creating ambient awareness through rich media links as being more concerned with social interaction than with the coordination of work, pointing out that while the personnel involved in media space trials considered their technologies to be a success, technologies of this type have not been widely implemented. Gaver (2002) also listed problems with media-spaces, such as difficulties in choosing how much ‘background’ to include in a video feed.

Gutwin and Greenberg (2002) defined workspace awareness (one of several possible kinds of awareness) as “the up-to-the-moment understanding of another person’s interaction with the shared workspace” (p. 417). They presented a framework for understanding and designing workspace interaction which comprises three dimensions: What information makes up workspace awareness? (the who, what, where, when and how of collaborative work); How is workspace awareness information gathered? (consequential communication from bodies, feedthrough from artefacts, and intentional communication through speech and gesture); How is workspace awareness used in collaboration? (management of coupling, simplification of verbal communication, coordination, anticipation, and assistance). Intentional and consequential communication correspond roughly to Goffman’s information ‘given’ and ‘given off’.

Gaver (2002) described technologies that extended the concept of awareness beyond the workplace and beyond the senses of hearing and vision. The “Feather”, “Scent” and “Shakers” systems were designed to convey emotion between separated loved ones, and used olfactory and tactile media. Such systems were usually not tested outside the lab: Gaver pointed out problems that might occur in real-world use due to the constrained communication these devices enabled. The “Bench Object” was designed to bring together strangers who happened
to be temporarily co-located in public settings. The “Schedule” and “Projected Realities” projects used visual aesthetics to create ambient cultural messages.

CSCW researchers have subsequently explored the concept of awareness in such diverse contexts such as hospitals (Bardram and Hansen, 2010), home-based aged care (Riche and Mackay, 2010), and software development teams (Omoronyia et al., 2009).

In their preface to a collection of recent work on awareness, Markopoulos et al. (2009) argued for a broader perspective on awareness that takes into account that “awareness systems have moved from the office into domestic and health-care environments and are starting to appear on mobile devices as well” (ibid., p. vii). Outside the work context, people use social networking systems (SNS) to maintain awareness of friends, not merely to efficiently perform tasks but to maintain social ties. Many of the messages exchanged for the purpose of maintaining social relationships are not informational but phatic (Vetere et al., 2009). And many people seek this awareness through the use of SNS while at work.

Markopoulos and colleagues argued that CSCW’s task-focused view of awareness be extended in time (to periods of months or years), and that “complete awareness of others” not be seen as an aim to aspire to: instead people want to control both transmission and reception of awareness information.

The application of awareness within collaborative virtual environments is discussed below in section 2.3.

**Social Translucence**

An influential framework for understanding awareness in CSCW systems is Social Translucence (Erickson and Kellogg, 2000). Social translucence has two dimensions. First, socially translucent systems make socially salient information available to participants, allowing them to be aware of others and their actions. This awareness helps bring the social rules, norms and customs that govern participants’ actions and interactions into effect. It also allows participants to be held accountable for their actions. Visibility, awareness, and accountability are crucial to the sociability of digital environments because they enable participants to structure their interactions with one another in coherent and sensible ways (Erickson and Kellogg, 2000).

Second, socially translucent systems are not transparent; rather, communication within these systems is constrained. In the physical world, physical laws constrain communication and determine who can receive what information in what circumstances. People use these
constraints as a resource in social interaction. For example, it is far easier to see and hear a person standing by one’s side than it is to see and hear someone across a crowded room, and this proximity constraint is routinely used in actions such as raising or lowering one's voice depending on one’s intended audience and social situation. Similarly, in designing socially translucent systems, attention needs to be paid to the design of the ‘physics’ or rules that govern communication within these digital environments. Furthermore, if participants share an awareness of the constraints that underlie the visibility of socially significant information in a digital environment they can use this understanding as a resource for structuring their social interactions.

These principles can be applied to the design of sociable online games (Gibbs et al., 2004). It is important for users to be able to easily discern and identify the other users with whom they are interacting, and to be able to readily associate game activity and actions with particular identities and personas.

Section 2.3 has reviewed research into mediated communication, focusing on the influence of modality on communication, the debate around media richness, and the importance of context. Section 2.2 reviewed research into virtual worlds. Now in section 2.4, I combine these themes by reviewing prior work on communication within virtual environments.

2.4 Voice in virtual space

In his preface to “The Social Life of Avatars”, Schroeder (2002) wrote:

The study of VEs not only benefits from comparisons with other forms of CMC, but also from comparing different modalities inside VEs. (p. 6)

A number of researchers made such comparisons. Sallnas (2002) conducted experiments to compare the influence of modality in Active Worlds. In one task, pairs of users who were unknown to each other made joint decisions about a fictitious scenario, communicating by text, voice, or a video link. Pairs made faster decisions and reported greater social co-presence when using voice or video instead of text. Video added no advantage over voice and may have detracted from efficiency.

Huxor (2001) conducted participant observation of a virtual office in Active Worlds, communicating by text. He found that after initial pleasantries on meeting, users usually
phoned each other for in-depth discussion, as they found conversing by text to be too cumbersome.

Nilsson et al. (2002) supplemented *Active Worlds* with a shared-audio system and studied workplace meetings held in the virtual world. They observed problems with turn-taking in the audio channel, concluding that spoken conversation did not work as well as it did in the physical world, especially for deictic reference to objects. The biggest problem encountered by users was malfunctioning audio. The text channel was used as a backup to the voice channel, and to clarify utterances or spell out a word. Some participants felt that text gave a better sense of presence than voice, possibly because it was more reliable, or more associated with computer use.

Slater and Steed (2002) arranged for a group of actors to conduct rehearsals for a play using a voice-equipped ‘desktop’ CVE. They noted that: “it is essential that actors communicate with one another with their full armoury of emotional expression and non-verbal behaviour” (p 157). The experiments were regarded by the users as reasonably successful. Problems arose from the need for the actors to manually control their avatars’ movements and gestures.

However as research into communication media (reviewed in section 2.3) has shown, voice projects information such as a user's gender, ethnicity and age, interfering with the anonymity which many VW users value (Raybourn, 2001) and which is exploited when role-playing. Bartle (2003b) argued that voice communication is incompatible with role-play and thereby breaks one of the fundamental purposes for being in the VW.

Koivisto (2003) described verbal and non-verbal communication in VWs, the latter including “character proximity, clothing, house decorating, possible emotes, or actions such as killing”, and indirect communication through changing the state of the virtual world.

Seay et al. (2004) examined tools used by MMOG guilds for team management. These included instant messaging systems, bulletin boards, reputation systems, ways for users to find and join guilds, and text chat within the guild. They identified problems such as ‘battle spam’, where a large number of game status messages interfere with personal communication.

Seay et al.’s informants felt that voice would improve their performance, but they did not use third-party tools, because of the administration overhead. Of the respondents, 78% were in guilds with an average size of more than 40. Respondents reported communicating for support and advice (77%), social exchanges (77%), and coordination and scheduling of activities (76%). Overall, 69% of players communicated outside the game via web-based message boards, instant messaging and email.
My colleagues and I studied the use of voice in *Xbox Live*, a large Internet-based network of game consoles which at the time of the study offered mostly first-person shooters and racing games (Gibbs et al., 2004). This network placed ad-hoc groups of users into a game and provided them with a shared voice channel for the duration of the game. We found that while users expected voice to be straightforwardly advantageous, the channel had usability and sociability problems. Users often couldn’t match voices to avatars or on-screen names. They were concerned that they did not know, and could not control, who could hear their utterances. They had no indication of whether their utterances could be heard, and often had to ask “is anybody there?” or “can anyone hear me?”. The voice channel was frequently abused through deliberate transmission of music or noise. However users appreciated the advantages of voice for coordinating fast-paced team games, as well as the sociability that voice added, which made Internet-based games feel more like a LAN party. We found that voice connected local and online social spaces, because microphones transmitted not only the utterances of their user but of other people co-located with the user.

Williams et al. (2007) provided voice to small pre-existing groups in *World of Warcraft* for one month and measured its effect on participants’ social attitudes and relationships, using a questionnaire. They found that team-mates who communicated by both voice and text during the trial liked and trusted each other more, and became happier and less lonely, than those who communicated only by text. However it was uncertain in this study how the participants used voice in different gameplay contexts, how often they reverted to text, whether they ever experienced problems with voice, how it affected their success in the game, or whether different user types had different media preferences. The authors did not test the efficacy of voice for large groups, nor for users who did not already know each other.

Erickson et al. (2011) studied a large, three-day conference held by IBM in *Second Life*. Four hundred IBM technologists from around the world logged in from their homes or offices and attended a mix of keynotes, poster sessions and socials, designed to approximate the social atmosphere of a face-to-face conference (and save travel costs in the post-2008 financial downturn). Although most attendees felt the event was successful, they experienced many technical problems with avatars and the spatial voice system. Some complained that *Second Life* avatars looked younger, fitter and more casually-dressed than a typical executive, thus reducing realism and making it hard to recognize people. Voices carried further than many expected, so that people became part of conversations distant to them. Being unable to manage small-group conversations caused sufficient discomfort that some individuals didn’t use voice.
These authors noted that attendees at the conference were simultaneously seated at computers in their physical-world homes and offices where they could be interrupted. This simultaneous occupation of two social spaces enabled forms of interruption (negative) and multi-tasking (positive) that the face-to-face situation did not afford.

A number of researchers and educators have embraced Second Life’s voice facility for language teaching. For example Jauregi et al. (2011) paired foreign language students with native speaker tutors and had them explore the virtual world and conduct tasks together. Participants praised the combination of simulated environment, pseudonymity through avatars, and rich interaction through voice.

In this section I have reviewed prior work on the use of voice in virtual worlds. My own approach to studying this phenomenon was informed by the recent move in HCI research towards studying user experience. This field of research is reviewed in section 2.5.

2.5 User experience

The User Experience (UX) movement within HCI has sought to broaden the scope of research into technology use, away from a focus on usability and the successful achievement of tasks and towards a broader consideration of aspects of experience such as context, emotion and enjoyment. McCarthy and Wright (2004) argued that:

It is no longer considered sufficient to produce a computer system that is effective, flexible, learnable, and satisfying to use … it must now also be useful in the lives of those using it. (p. 5)

In a review of UX studies published between 2005 and 2009, Bargas-Avila and Hornbaek (2011) detected:

… a shift in the products and use contexts that are studied, from work towards leisure, from controlled tasks towards open use situations, and from desktop computing towards consumer products and art.

These emphases on recreational technologies, on context, and on enjoyment suggest that UX is an ideal framework for researching use of virtual worlds.

UX takes a holistic view of the experience of using a technology, aiming to understand how the characteristics of users, products and social contexts together influence experience. Law et al. (2009) remarked:
HCI researchers and practitioners have become well aware of the limitations of the traditional usability framework, which focuses primarily on user cognition and user performance in human-technology interactions. In contrast, UX highlights non-utility aspects of such interactions, shifting the focus to user affect, sensation, and the meaning as well as value of such interactions in everyday life. (p. 719)

The emphasis on context of use has caused the locus of study to shift from the laboratory into homes, workplaces and mobile situations. McCarthy and Wright (2004) argued that technologies and technology use are embedded in everyday life. People appropriate technologies and make their presence and use routine. While use of a technology can therefore be mundane - in the sense of ‘everyday’ - it is by no means precluded from provoking intense feelings.

Our interactions with technology can involve emotions, values, ideals, intentions, and strong feelings (McCarthy and Wright, 2004, p. 2)

Technology use is rarely a solitary pursuit. Forlizzi and Battarbee (2004) advocated an interaction-centred framework for understanding user experience, including not only user-product interaction but user-user interaction around a product. They described interactions such as occur around museum exhibits and through messaging systems as ‘co-experience’ – the creation of meaning and emotion together through product use.

Co-experience reveals how the experiences an individual has and the interpretations that are made of them are influenced by the physical or virtual presence of others. (Forlizzi and Battarbee, 2004, p. 263).

This concords with the concept of co-presence in CVE research.

The UX approach is well-suited to recreational contexts and has become important within HCI during the period that technologies such as smartphones, social network sites and computer games have become mass-market products. UX offers a way to understand consumer use of non-workplace technologies in non-workplace contexts that usability testing cannot capture.

It is as much about children playing with GameBoys, teenagers gender swapping, and elderly people socializing on the Internet as it is about middle-aged executives managing knowledge assets, office workers making photocopies, or ambulance controllers dispatching ambulances. (McCarthy and Wright 2004, p. 9)

Experience is of course profoundly subjective. As a result, the user experience movement has included a shift away from quantitative research methods. Bargas-Avila and Hornbaek (2011) found that while only 14% of usability studies were qualitative, over half of published UX research had embraced qualitative methods. These are more able to capture the richness and complexity of user experience, and are more open to unexpected findings. Bargas-Avila and Hornbaek quote Swallow et al. (2005) on the inapplicability of quantitative methods to capturing experience:
A common strategy ... is the reduction of experience into a number of factors or processes ... such approaches may be useful for experimental analysis but they can miss some of the insights available in accounts that resist such reduction ... qualitative data provides a richness and detail that may be absent from quantitative measures. (p. 91-92)

UX however is a nascent field whose “definition and distinct characteristics as a research field are currently unclear” (Bargas-Avila and Hornbaek, 2011, p. 2690). Because it is not clearly defined, UX can mean different things to many people, and there is no agreed-upon set of methods for studying it. McCarthy and Wright (2004) caution that the term ‘user experience’ has been used as a marketing buzzword. In their review of UX research, Bargas-Avila and Hornbaek discovered a number of criteria on which trends could be detected but on which the UX community did not have consensus. These included:

1. Products, use situations and context: 64% of studies assessed leisure products rather than work products. 61% assessed open use or left the decision to use up to participants, rather than assessing controlled use such as lab tasks. Half controlled the context of use (e.g. in a lab) while a third used uncontrolled settings such as online use.

2. The dimensions of experience measured: 41% sought to assess the overall experience while the remainder assessed a specific dimension of experience. Of those studying overall experience, 80% were qualitative studies. Where specific dimensions of experience were assessed, the most common were emotions and affect, enjoyment and aesthetics.

3. Data-collection methods: Half of UX studies were qualitative, a third quantitative and the remainder used both approaches. Most of the quantitative studies used questionnaires. A wide range of qualitative methods were used, including ‘probe’ methods. (Bargas-Avila and Hornbaek, 2011)

Bargas-Avila and Hornbaek reported that the plainest dichotomy was between the ‘general experience’ studies, usually using qualitative methods, and the ‘specific dimension’ studies, using mostly quantitative methods, suggesting that two research cultures are using the UX name but have different understandings of what experience is and how it should be studied.

While Bargas-Avila and Hornbaek emphasized the lack of consensus among UX researchers, this can partly be attributed too the immaturity of the field. Most authors acknowledge an overall focus on experience over usability, on the uncontrolled and unpredictable social context of everyday life, and on recreational or ‘lifestyle’ use of technology over the performance of tasks in a work setting.
2.6 Gap

My review of prior work demonstrates that while a significant amount of research has been conducted into the use of virtual worlds, and into mediated communication, comparatively few studies have specifically addressed the junction of these two fields: that is, mediated communication within virtual environments.

Many of the studies that addressed this intersection utilized early experimental CVEs, which did not have some of the key properties of current virtual worlds such as a large user-base, interaction among strangers and role-play in fictional settings.

Data-driven research on VWs has usually examined text communication only (Harris et al, 2009 is an example) possibly because text utterances are easier to store and count than voice utterances. Only one study (Williams et al., 2007) compared modalities within a modern virtual world, and it was limited to one system, a game (World of Warcraft), and to small groups whose members knew each other, both of which are factors likely to influence the suitability of voice.

The UX approach, with its emphasis on context, emotion, the holistic experience of using a technology and the influence that technology has in people’s lives, has rarely been brought to bear on research into virtual worlds.

This is the gap I have attempted to address.

At the start of my research, anecdotal evidence and prior work on multiplayer games indicated that voice was having a significant influence on the user experience of virtual worlds. I perceived a need to examine the influence of the introduction of voice across a range of virtual worlds, users, uses and contexts. I felt this influence was likely to be complex, and that in order to understand it I should gather rich data on the experiences of the people using these technologies.

2.7 Research question

Given the gap in prior research identified above, I chose as the overall research question guiding this thesis:

RQ: How does voice influence the user experience of virtual worlds?
Because the use of voice in virtual worlds had been little examined, I needed to conduct exploratory research and collect data in such a way as to enable the generation of new theory. Thus I adopted a grounded theory approach to choosing cases and to collecting and analysing data. My approach is discussed in chapter 3.

To address the wide diversity of types of systems, users and contexts, I conducted four case studies. Collectively, these address different kinds of communication systems in different kinds of virtual world. The studies are described in detail in chapters 4 through 7.

I refined my research question to produce individual questions to be answered by each of the case studies.

Case 1: The introduction of voice to massively-multiplayer online games.

RQ1: How does voice influence the user experience of massively multiplayer online games?

Case 2: Spatially-propagated voice in virtual environments

RQ2: How does spatially-propagated voice influence social interaction and tactics at the ‘macro’ level of spatiality in multiplayer videogames?

Case 3: Communication during collaborative manipulation of objects in a virtual world.

RQ3: How do groups communicate using voice while collaborating around objects at the ‘micro’ level of spatiality in a virtual world?

Case 4: The introduction of voice communication to a social virtual world.

RQ4: How did the introduction of voice influence the experience of different kinds of users in Second Life?
2.8 Conclusion

In this chapter I have reviewed relevant prior research into fields of virtual worlds, mediated communication and user experience.

I noted where the phenomena studied in these fields were likely to interact: for example the richness of a communication medium affects its ability so support presentation management, which is likely to impact role-play when the medium is adopted by virtual world users.

I established that few studies have directly addressed mediated communication within virtual worlds, so that the influence of modality upon the user experience of VWs is largely unexamined. Prior research on mediated communication more generally suggests that modality should have a significant and complex influence on user experience.

In the following chapter I describe and justify the approach I took to address this gap in knowledge.
Chapter 3

3. Design of my research program

3.1 Introduction

In the previous chapter I reviewed prior research into virtual worlds, mediated communication and user experience. I explained that my research was located at the intersection of these three fields and established that little prior work had been conducted to understand how communication modality influences user-experience within virtual worlds. This gap in knowledge led me to choose the research question:

*How does voice influence the user experience of virtual worlds?*

In this chapter I describe and justify the design of the research program that I devised to answer this question. I begin by analysing the nature of the phenomena that my research was intended to examine.

3.2 Phenomena and challenges

3.2.1 Virtual worlds

Schroeder (2002) argued that a fundamental problem encountered when studying social interaction in virtual worlds is understanding exactly where the interaction is happening, and by extension, where and how data collection should occur. All technology use happens within a context, and research reviewed in section 2.2 showed that context is an important influence on use. Context includes (but is not restricted to) who is doing what in the place in which use occurs (Dourish, 2004). Yet virtual worlds *simulate* a kind of place themselves. So the VW user has two kinds of context – physical and virtual - to be considered when investigating influences on his experience. For example Moore et al. (2007) argued that a system could
conceivably deliver awareness about both physical and virtual contexts in order to allow users to avoid interrupting discussions of collaborators in either context.

Some virtual worlds researchers have taken an ethnographic approach and chosen to locate themselves as observers in the virtual space, interacting only with users’ onscreen representations. For example Boellstorff (2008) conducted an ethnography of Second Life in which he chose not to interact with users face-to-face, but treated them as entities that existed entirely within the virtual world. This approach has been called ‘virtual ethnography’ (e.g. Moore et al., 2007). It makes the object of its study not the users of a VW but rather the personae they project into it.

However the approach of regarding ‘real world’ and ‘virtual world’ as separate realms has been criticized by Lehdonvirta (2010), who argued that:

The roots of this dichotomy can be traced to the magic circle concept in game studies and the cyberspace separatism of early Internet thought. … The world of players does not respect the boundaries of an MMO server, as it frequently flows over to other sites and forums. At the same time, other social worlds, such as families and workplaces, penetrate the site of the MMO and are permanently tangled with the players’ world. Research programs that approach MMOs as independent mini-societies are therefore flawed, but there are many other kinds of research that are quite feasible. (online)

Schroeder (2010, pp. 213-4) has also criticized the approach of disregarding the offline actions and characteristics of VW users. And Golub’s (2010) argument against the importance to MMOG players of visual realism helps to reject the supposition that these users are immersed in a separate realm unconnected to the real world.

In a similar vein, the research into mediated communication reviewed in the last chapter showed that choice of modality influences and is influenced by the context of a conversation. To pick two examples, mobile phone users choose text messaging to evade eavesdroppers (Madell and Muncer, 2007), and the voice channel in Xbox Live connects the physical and virtual social contexts of its users (Gibbs et al, 2004).

Therefore I did not regard virtual worlds as separate realms, but rather as technologies used by people in real physical and social contexts. I argue that to understand what it is like to use a virtual world requires taking into account the user’s experience of simultaneously occupying points in physical and virtual space. Researchers must seek to understand the circumstances in which events in either of these two spaces can influence events in the other.

Note that taking this approach does not preclude consideration of the ‘immersionist’ style of use. The fact that a user may (wish to) feel immersed in a virtual world certainly does not preclude them from being influenced by their physical surroundings – if anything it can render
physical-world influences more problematic and therefore more important. I wanted to understand how the characteristics of users and contexts influence choice and use of modality, and how in turn the characteristics of the modality used influence the experience of using virtual worlds.

When studying virtual worlds, another consideration is sampling: the set of systems is large and always changing as new systems are released and old ones go out of production. The problem of how to address this diversity led me to study multiple scenarios in multiple VWs. Yet understanding UX requires the investigation of scenarios in depth, which restricted the number of cases I could study. This dilemma necessitated the choice of a small, representative sample.

Though the number of virtual worlds is large, the population can be reduced by sampling from categories. A taxonomy I described in chapters one and two divides VWs into two types, game and non-game worlds – in other words MMOGs and social or open-ended worlds. Use of virtual worlds can be classified as either recreational or work-oriented, and approaches to use can be classified as either immersionist or augmentationist. I selected cases in such a way as to maximize the comparisons across these types.

Another question was how to organize the research in time. Each existing virtual world had, until a particular point in time, offered only text communication. Then voice became available, sometimes first through users' appropriation of third-party systems and later through developers' integration of voice into the VW in question. For a given VW therefore, one can speak of 'before and after voice', and ask what influence the introduction of voice had. This implies a comparison of use at times before and after the appearance of voice, which in turn suggests a longitudinal or repeated-measures approach, gathering data first in the text-only era and later in the voice era. By holding the system and participants constant, such an approach would minimize confounding factors. On the other hand it would be impossible (or at least non-naturalistic) to use a counter-balanced design in which the order of treatments was randomized, because text always comes before voice. Longitudinal studies require a greater commitment by participants, and holding the study system constant would mean that, for a given total number of cases, fewer types of virtual world could be examined.

Instead I chose to conduct a series of ‘cross-sectional’ studies (Neuman, 2011), examining the experience of users in particular (kinds of) virtual worlds at particular points in time. This avoided the constraints of longitudinal studies and offered the freedom of working within a broader range of use scenarios. The cost of this approach was that the influence of voice was necessarily examined at points after it was introduced: for many participants, using a text-only...
VW was a mere memory. Ameliorating this somewhat was the fact that data-gathering for my final study took place over a relatively long timeframe and began shortly after the introduction of voice into the study system.

### 3.2.2 Methods used to study virtual worlds

A number of methods have been employed by researchers studying use of virtual worlds. Sometimes the focus has been on usability as understood traditionally within HCI. Testing usability can be important (e.g. Schroeder et al., 2006), but to capture what happens in large multi-player systems requires a richer set of methods, such as those used in CSCW research and ‘third paradigm’ HCI. In large systems the interactions between users become more worthy of study than the interactions between user and computer.

Schroeder et al. (2006) discussed the appropriation of methods from ethnography, social interactionism and conversation analysis to study virtual environments, noting that while these produce rich understandings of particular situations, the results can be difficult to generalize to other settings.

A consideration for choice of method is whether the researcher is able to control the setting and participants. In the experimental CVEs of the 1990s, all of the users of a system were usually known to the experimenter, and settings could be constructed and controlled. By contrast, when studying the large, commercial, Internet-based worlds that have appeared recently, a researcher typically has no special access to the system but is simply one of the users. In some studies, the participants are users encountered randomly in the VW, so that the practical and ethical problems of Internet research methods apply.

Ethnographic methods such as observation, participant observation and interviews have been popular for research into modern virtual worlds. Researchers have conducted ethnographies of MMOGs (Taylor, 2006; Nardi and Harris, 2006; Ducheneaut et al., 2006; Bainbridge, 2010; Nardi, 2010), as well as “social” worlds (Taylor, 1999; Boellstorff, 2008). Some have conducted ethnography which takes into account the users’ offline lives (e.g. Taylor, 2006). Others have preferred to conduct ‘virtual ethnography’, studying only the personas projected into the VW (e.g. Moore et al., 2007; Boellstorff, 2008).

Some researchers have conducted quantitative, data-driven research. To gather large amounts of usage data within a public system with which the researcher has no special relationship requires some ingenuity. One approach is to survey users. For example, Yee (e.g. 2006) was
able to recruit large numbers of MMOG players for a survey. Many have surveyed *Second Life* users, typically with smaller samples. Another form of quantitative analysis available in virtual worlds is the programmatic measurement of behaviour. Yee et al. (2007) studied avatar proxemics in *Second Life* using software that recorded the locations, orientations and names of nearby avatars. The software was operated by research assistants, who manually noted the gender of the avatars (which could not be detected by the software). In this way a large amount of data was recorded for statistical analysis.

Other studies have combined qualitative and quantitative methods, seeking to exploit the deep understanding attainable through ethnography with the generalizability of surveys. Williams et al. (2006) argued for the benefits of this hybrid approach. This group took a ‘census’ of an MMOG by setting up bots which over time regularly extracted data about currently logged-in users including their guild membership. They combined this method with interviews of users to develop an understanding of guild dynamics.

Recently a group led by Williams was granted access to the database of the MMOG *Everquest 2*. This has enabled quantitative studies of essentially the entire user-base – in contrast to the more usual sampling approaches. A recent publication from this project is by Wang et al. (2011). These authors matched log data with a survey of 7000 players to explore relationships between players types and usage patterns.

Some researchers have employed conversation analysis (CA) to study interaction in virtual worlds. (I used a variant of this method in study 3.) Bowers et al. (1996) used CA to study a project meeting held in one of the first networked virtual worlds, *MASSIVE*. Hindmarsh et al. (2000) used CA to study object-focused collaboration in the same system. Schroeder et al. (2006) compared two methods, CA and a quantitative categorizing method, in the context of research into CVE usability. Brown and Bell (2004) studied the virtual world *There* using a combination of virtual ethnography and interaction analysis. Moore et al. (2007) and Ducheneaut et al. (2007) video-recorded encounters in MMOGs and used CA to analyse ‘virtual face-to-face’ interaction. (These studies are described in more detail in section 2.2.12.)

A conversation analyst audio- or video-records talk ‘in the wild’ and transcribes it into text for analysis. The layout of the transcript is designed to reflect the ordering of utterances, as well as any overlapping of utterances, and any silences between or within utterances. Transcription conventions include special characters for representing short silences and the length of longer ones, and precisely where overlaps occur. Attention is paid to sounds such as ‘oh’ and ‘uh’, which interlocutors use to effect turn-taking and (self-) repair of interaction (Ten Have, 1999).
Description of actions other than speech can be added as required. In Jefferson’s transcription convention:

The basic technique is to visualize on paper the timeline of the interactional stream, and to place each participant’s contribution in relation to others. The space occupied by the letters in the printed words is taken as a visual image of the length it took to produce the corresponding sounds during the interaction. And the details added to the textual transcript serve to make this picture more complete and exact, noting pauses, overlapping, slower, faster, latched or stretched speech. (Ten Have, 1999, p. 32).

CSCW researchers who use CA have varied in the extent to which they use these typographical conventions. Some have omitted markup in favour of including ethnographic description which explains what is occurring while participants are talking. Take for example this fragment from Hindmarsh et al.’s (2000) study of collaboration in a CVE:

Example 6: A30/1/98-10:31:50-VP:M.
M: Do you want the -er lamp moving?
T: Could do. see what it looks like.
M: Where d’ya want it put?
T: Erm- you decide
((pause as M moves the lamp))
T: That’s right in front of the window, that’s probably not so clever actually. over the other side, over here. (p. 495)

This conversation is presented with little markup but is followed by a paragraph explaining what the users were doing and what went wrong. This is the presentation style I have used for study 3.

3.2.3 User experience

A major consideration for this thesis was how to collect data about people’s experiences. UX research seeks to understand the holistic experience of people interacting with and through technology; this requires the researcher to go beyond concerns with usability and efficiency and study meaning and emotion.

UX emphasizes the role of the technology in daily life (Law et al., 2009; McCarthy and Wright, 2004) and therefore the influence of context, including social context. Experience is “influenced by the physical or virtual presence of others” (Forlizzi and Battarbee, 2004), and “recognizes that computers are employed within situations of human interaction and collaboration” (Crabtree, 2003). In other words, to study UX one must acquire a rich understanding of the total influence that a technology has on a person’s day-to-day life.
In their review of UX research, Bargas-Avila and Hornbaek (2011) noted that “a key issue is whether UX may be measured and modelled, and if so how”. These authors discovered that half of UX studies were qualitative, relative to only 14% of HCI studies in general. Far more leisure than work products were studied by UX researchers, and only a third of UX studies used controlled-task scenarios, the majority employing open-use or user-initiated-use scenarios – though half used non-naturalistic settings such as labs.

The trend toward a broad understanding of experience aligns with the “third paradigm” that Harrison et al. (2007) argued has recently gained traction in HCI research. This post-cognitive, phenomenological approach embraces the total experience of the user and allows characteristics of users and elements of context that are not directly related to the technology interaction to be detected if they are important to the user’s experience. These elements might have been ignored or factored out as noise in earlier usability-focused research.

Meaning … is, instead, irreducibly connected to the viewpoints, interactions, histories, and local resources available to those making sense of the interface and therefore to some extent beyond the reach of formalization. (Harrison et al., 2007, p. 7)

Bargas-Avila and Hornbaek (2011) found that 40% of UX studies (generally the qualitative ones) sought to understand what these authors called ‘generic user experience’. (Many of the remainder sought to measure a specific emotional response such as enjoyment.) The qualitative studies used a variety of methods which:

... involve talking directly with users, such as semi-structured interviews (20%), focus groups (15%) and open interviews (12%), as well as less intensive procedures such as user observation (17%), analysis of videorecordings (17%) and diaries (11%)” Bargas-Avila and Hornbaek (2011, p. 2693).

My approach to research is aligned with this. I have used qualitative methods such as open-ended interviews and focus groups to try to understand the complete experience of people using virtual worlds, and what influence the introduction of voice has had.

However in study 3, where conversation around objects was the focus, I also incorporated a variant of conversation analysis, as described in section 3.2.2.

3.3 Overall approach

The user experience literature suggested that I should gather rich data about subjective experience and co-experience in naturalistic settings. The literature on mediated communication suggested that the influence of modality was likely to be complex and
personality- and situation-dependent. Prior research into *virtual worlds* emphasized the diversity of VW systems, users and uses, and has typically used ethnography or interaction analysis or both.

Together these considerations guided me to take an exploratory, grounded theory approach, and employ primarily the qualitative methods preferred in UX and much VW research (Bargas-Avila and Hornbaek, 2011), supplemented by detailed interaction analysis and quantitative measures as appropriate.

I designed a research program consisting of four cross-sectional case studies corresponding to important use scenarios. Swallow et al. (2005) noted that:

One of the most useful qualitative research methods available is the case study, which focuses upon only one or a few instances of the population or situation under investigation. Advocates of this approach argue that while statistical methods might be able to deal with situations where behaviour is homogeneous and routine, case studies are necessary in order to account for creativity, innovation, and context.

I can describe my program succinctly using Neuman’s (2011, pp. 26-53) five dimensions for analysing research projects.

1. **Basic vs Applied:** Voice in virtual worlds had been little studied, so my project was basic research aimed at generating theory. However I intended also that my results should help inform design.

2. **Exploratory, descriptive or explanatory:** I began with exploratory studies; later shifting emphasis to descriptive and explanatory studies.

3. **Within or across cases:** I conducted multiple case studies.

4. **Single or multiple points in time:** Each individual case study was cross-sectional rather than longitudinal.

5. **Quantitative or qualitative:** I gathered predominantly qualitative data, but performed some quantitative analysis in study 3.

The remainder of this section describes and justifies my broad methodological commitments in greater detail.

### 3.3.1 Exploratory

Because voice communication was a new phenomenon in online virtual worlds, without a body of research to shed light on it, my research was necessarily to a degree exploratory.
Stebbins (2001) passionately proclaimed the necessity for exploratory social science research, describing this approach as:

… a broad-ranging, purposive, systematic, pre-arranged undertaking designed to maximize the discovery of generalizations leading to description and understanding of an area of social or psychological life. (p. 3)

Stebbins argued that inductive exploratory research was necessary for social science to expand its boundaries, because deductive reasoning from existing theory cannot address new phenomena. He emphasized that exploratory research is systematic, and contrasted it with serendipity or accidental discovery.

Stebbins recommended qualitative methods but emphasized that ‘exploratory’ and ‘qualitative’ are not synonyms. Exploratory research is concerned not just with collecting data but with developing theory from it. For this reason Stebbins aligned exploratory research with Glaser and Strauss’s (1967) grounded theory methodology.

### 3.3.2 Generating theory from data

Although I was interested in the experiences of people using virtual worlds, I did not merely wish to describe these experiences, but rather sought to try to explain the relationship of experience to other phenomena such as modality. That is, I wanted to generate theory. As a result I chose the grounded theory approach (Glaser and Strauss, 1967; Strauss and Corbin, 1998) to inform my overall approach.

Although a phenomenology emphasizes the meaning of an experience for a number of individuals, the intent of a grounded theory study is to move beyond description and to generate or discover a theory, an abstract analytical schema of a process. (Creswell, 2007, p. 63).

Grounded theory (GT) seeks to generate theory systematically from empirical data. It favours the discovery of new theory over verification of existing theory.

A researcher does not begin a project with a preconceived theory in mind … Rather, the researcher begins with an area of study and allows the theory to emerge from the data. (Strauss and Corbin, 1998, p. 12).

This emphasis on gathering data in order to create new theory, rather than fitting pre-existing theories to new situations, makes GT particularly suited to exploratory research.

Grounded theory is a good design to use when a theory is not available to explain a process. The literature may have models available, but they were developed and tested on samples and populations other than those of interest to the qualitative researcher. … A theory may be needed to explain how people are experiencing a phenomenon, and the grounded theory developed by the researcher will provide such a general framework. (Creswell, 2007, p. 66).
GT guided my selection of cases and my collection and analysis of data. Often qualitative data is gathered in naturalistic settings. However GT does not exclude quantitative methods (or indeed any method), and embraces empiricism:

There is no fundamental clash between the purposes and capacities of qualitative and quantitative methods or data. (Glaser and Strauss, 1967, p. 17)

In many instances, both forms of data are necessary – not quantitative to test qualitative, but both used as supplements, as mutual verification and, most important for us, as different forms of data on the same subject, which, when compared, will each generate theory. (ibid., p. 18)

GT encourages the researcher to remain open to unexpected findings and to change direction when results demand. The GT researcher uses the principle of theoretical sampling to select new sources of data as a project proceeds. Accordingly I used the findings of early studies to inform the choice and design of later ones, so that my program broadened in directions suggested by the emerging theory. My focus shifted from theory-building to theory-testing, but I remained open to new insights and these occurred even in the final study.

The principle of comparative analysis which is at the heart of GT is visible in my research design. As evidenced below in section 3.4, my set of studies was chosen to enable comparison along three dimensions important to my topic: type of virtual world, usage style and spatial scale. Comparison was also part of the design of individual studies. Study 1 was based on a comparison between two ways of implementing voice systems, and between groups who knew each other and groups who did not. Study 2 compared spatial with radio voice. Study 3 compared use in a modern virtual world with results from earlier CVEs, and involved comparisons between skill levels and task types. Study 4 drew data from people who used Second Life for very different purposes.

The grounded theory approach is popular in HCI research, though different researchers adopt different variants and there has been debate over how the methodology should be used in this field. Problems can arise that relate to timing constraints and to the use of a literature review and prior theory, so that researchers tend to blend the influence of institutional requirements with grounded theory methods (Furniss et al., 2011).

3.3.3 Ethnography

Researching technology use in naturalistic settings suggests the employment of ethnographic methods such as observation, interviews and diaries. These are collectively termed ‘field research’ by Neuman (2011).
As noted above, one difficulty in conducting naturalistic research with virtual world users is that the action to be investigated takes place both in a physical and a virtual context. This raises practical difficulties for data collection, especially for observation methods. In only two studies (2 and 3) did I directly observe users in their offline context. In another (1), use occurred in natural settings, but data collection was via diaries and interviews made after the action was over.

Creswell (2007, p. 68) differentiates the grounded theory and ethnographic approaches in that the latter “focuses on an entire cultural group”. Virtual worlds can be considered ‘places’ which have ‘cultures’, and several researchers have conducted ethnographies of these (Taylor, 2006; Bainbridge, 2010; Nardi and Harris, 2006; Ducheneaut et al., 2006; Taylor, 1999; Boellstorff, 2008). In studies 3 and 4, which utilized Second Life, I took such an approach - that is, “long-term, immersive field work combining observation with participation” (Dourish, 2006). However acknowledging this author’s admonishment of HCI researchers who appropriate ethnography’s methods without adopting fully the ethnographic stance, I do not claim to have written ethnography, but merely to have borrowed its methods.

### 3.3.4 Interpretation of experience

Much UX research makes use of interviews to collect subjective data (Bargas-Avila and Hornbaek, 2011). This requires that participants first interpret their own experience: a ‘first-order interpretation’ in the schema of Neuman (2011, p.177-8). Subsequently the researcher must interpret these data for himself (second-order), and finally interpret them in the light of theory (a third-order interpretation). Thus to understand the experience of others necessitates an interpretive approach to data analysis. Neuman (2011) describes the interpretive stance as: “the systematic analysis of socially meaningful action through the direct detailed observation of people in natural settings”.

My own stance as a researcher is to understand that when I ask a participant to describe and explain their experiences and what influence a particular technology had on them, I am asking them to interpret reality. Asking for an interpretation is not to deny that objective reality exists, but simply acknowledges that the complex social reality described by an interviewee is an interpretation of their subjective experience.

To gather and interpret subjective data is not to deny that the phenomenon of interest is objectively observable. I took opportunities in these studies to triangulate participants’
subjective interpretations against empirical observations. I take this approach to be compatible with ‘post-positivism’ as described by Creswell (2007, p. 20).

3.3.5 Mixed methods and triangulation

Triangulation seeks to increase confidence in research findings through employing multiple perspectives. Neuman (2011, pp. 164-5) lists four strategies: triangulation of measures, triangulation of observers, triangulation of theory and triangulation of method.

In my research, triangulation of measures occurred through the use of multiple case studies, multiple user groups within most studies, and case selection that maximized the diversity of the sample within the overall scope of the topic.

I used triangulation of method in my later studies. Study 3 involved qualitative and quantitative methods, and in study 4 I employed a range of qualitative methods.

The strategy of triangulation through multiple observers has limited availability to PhD researchers. However I was able to discuss interpretations with colleagues and supervisors. Each of my studies was presented at a conference (see appendix C), allowing me to discuss findings and interpretations with other researchers. I also reported findings back to user communities where possible, for example by posting publications to Second Life mailing lists.

Triangulation of theory is evident in my use of prior literature from multiple fields such as HCI, game studies, and communication research. An example concept important in this thesis which takes different forms in these different disciplines is context of use (Dourish, 2004), which is related to the ‘magic circle’ debate in game studies (Lehdonvirta, 2010) and control of visibility/audibility in communication research (Jones, 2002; Madell and Muncer, 2007).

In using both quantitative and qualitative data I embraced the mixed-methods approach as defined by Creswell and Plano Clark (2007), who argue that:

… the use of quantitative and qualitative approaches in combination provides a better understanding of research problems than either approach alone. (p. 5)

For these authors, mixed methods allows a researcher to use multiple worldviews or paradigms, and offsets the weaknesses of using either approach alone.
3.4 Selection of cases

I conducted four case studies:

Study 1: voice in massively multiplayer online games;

Study 2: spatial voice in a team combat game;

Study 3: communication during collaboration around objects;

Study 4: voice in a social world.

The selection of these cases was driven by the principle of theoretical sampling. Glaser and Strauss (1967) stated that:

The basic criterion governing the selection of comparison groups for discovering theory is their theoretical relevance for furthering the development of emerging categories. (p. 49)

Thus the first pair of studies were conducted in game worlds, because interesting appropriation of voice occurred first in this genre of virtual worlds. These studies surfaced relationships between modality preference and both game activity and context of use. It emerged that a paradigm situation to which voice was suited was coordination of teams in virtual space: a scenario that was important in case 1 and ubiquitous in case 2.

The theoretical sampling principle therefore suggested that after cases 1 and 2, I should proceed to study virtual worlds that did not feature this type of team activity, in this way expanding the range of observation along this dimension. Thus as a result of categories emerging in studies 1 and 2, I chose to study non-game VWs in cases 3 and 4.

The cases differed along another dimension as well. Recall from section 1.2 that a defining characteristic of virtual worlds is spatiality. Two of my four studies directly addressed this characteristic. In study 2 participants were exposed to a voice system that propagated sound according to users’ locations in virtual space. In study 3 participants were asked to perform tasks that involved communicating about objects and locations in virtual space. So where studies 1 and 4 were designed to draw out the general experience of using voice in a virtual world, and compare experiences between game and open-ended worlds, studies 2 and 3 focused on spatial experience.

Figure 5 illustrates the comparisons made possible by my selection of cases. The dimension of virtual world type runs top to bottom. The dimension of use of voice runs from left to right, and there is a third dimension that differentiated cases 2 and 3: spatial scale. Case 2 examined
action at the macro scale of spatiality – avatars moving through a large space - while case 3 examined the micro scale of avatars collaborating around object.

Another way to illustrate the relationship of the cases to each other is by placing them within the taxonomy of virtual environments that I illustrated earlier in figure 4, section 1.2. Figure 4 is adapted here as figure 6, in which the four studies appear as numbered circles according to test system.

This set of cases provides a diversity of contexts which Yin (2003) argues is necessary in order to isolate the unit of analysis (in this case, voice in virtual worlds) from the influence of context, to increase the likelihood of conclusions being generalizable beyond the cases examined.
3.5 Methods used in individual cases

I employed a variety of methods to gather data:

- individual interviews
- focus group interviews
- diaries
- observation
- participant observation
- interaction analysis of screen recordings
- analysis of user discussion forums.

The data gathered consisted of:

- Transcripts of interviews, focus-groups, diaries and forum posts,
- Notes made during observation, and
- Transcripts of screen recordings.

I coded the interview transcripts using the NVivo software package, to discover themes. Samples of data and worked examples of coding from each study can be inspected in Appendices D through G. In study 3, where I was able to make video recordings of use sessions, I conducted interaction analysis and made quantitative analyses of the recordings, counting “salient events” and comparing averages across cohorts of users. These are presented via charts in chapter 6.

Study 1: Voice in game worlds

My first study examined the influence of communicating by voice on the experience of using massively multiplayer online games. These complex games demand regular use over months or years and regular interaction with other users. Users become highly knowledgeable about the simulated world, its folklore and customs, and about their guild and their role in it. However many keep their offline identity private, and some engage in role-play and gender-
swapping. Given what was already known about mediated communication, it seemed likely that using voice would influence these complex social interactions.

There was little prior research on voice in virtual environments, and none on voice in modern virtual worlds. Therefore I adopted an exploratory approach (Stebbins, 2001). I considered observing MMOG players in a usability lab, but decided instead to try to capture the experiences of users in their natural playing environment. I used diaries as the data-gathering instruments as these are suited to recording experience in naturalistic research. I also employed intensive, one-on-one, open-ended interviews, and focus-group interviews.

Fifteen experienced MMOG players were recruited through personal and professional contacts. All had used voice in other types of games, so issues reported were due to the interaction of voice with the MMOG context. I formed them into three groups and asked them to use voice under different circumstances in different MMOGs. Two groups used a new MMOG which was the first to incorporate voice communication within the game software. Of these two groups, one consisted of friends who knew each other well and had played other games before. The second group of participants had never met each other before, and usually didn’t play together during the study but instead joined pickup-groups. This enabled a comparison between users who did and did not know each other offline, which prior research had indicated was a key issue for voice acceptance. A third group continued to use third-party voice products in the MMOGs they were already playing: this enabled a comparison of “integrated” and “third-party” voice.

To achieve naturalism, participants played in their own homes under their normal playing conditions. They kept diaries in which they recorded their experience of using voice in the game. Half way through the study, all participants were interviewed individually. At the end of the study, two of the groups participated in focus-group interviews. Interviews and focus-groups were open and semi-structured and lasted one to two hours. I sought to understand participants’ use of voice and to collect their criticisms of existing implementations and suggestions for future ones. I used ‘canned’ questions to seed discussions but allowed conversation to flow according to what users found interesting and wanted to talk about. The focus groups discussed six fictitious gameplay scenarios to consider how they could use existing or imagined voice systems to deal with each scenario.
Study 2: Spatial voice in team combat

The purpose of study 2 was to assess the influence of spatially-propagated voice. As was the case in study 1, no prior research existed into the use of voice systems of this type, and so again an exploratory approach was appropriate. Focus-group interviews and observation were the main instruments chosen.

The SpatialVoice product allowed a group of people playing a networked multiplayer videogame to speak with each other using microphones and headsets. It operated in two modes. One employed the familiar ‘radio’ metaphor for propagation: every member of a team was considered to be on the same channel, and all could hear all equally well. The other mode emulated the transmission of sound in air: the ability of users to hear each other depended on how close their avatars were in virtual space, regardless of whether they were on the same or the opposing team.

‘Radio’ mode is easier to implement, as the system need only know the IP addresses of each team member. By contrast, the spatial algorithm needs to know the locations of users’ avatars in virtual space, and the space’s topology and contents, in order to calculate at what volume each user should receive an utterance. This requires tight integration with the game software. At the time of the trial, SpatialVoice had been integrated with one game: a team FPS. In this the team members played different roles – gunner, engineer, medic etc – and to win it was vital to coordinate members’ actions.

A group of nine co-workers were observed playing a weekly lunchtime FPS game using SpatialVoice. Participants took part in focus groups after one month and three months of use.

Study 3: Communication during collaboration around objects

My third study took place at the Palo Alto Research Center (PARC), which I visited during my candidature. Here I conducted a study of how people communicate while collaborating around objects in a virtual world. I used Second Life because it was a popular system, offered object manipulation and had a vibrant user culture centred on building.

In this case, prior research existed on problems of collaboration in virtual environments. My goal was to update the understanding of this phenomenon by examining it in a modern system. I chose methods informed by the ‘furniture world’ research of Hindmarsh et al. (1998) and work by Fussell et al. (2003) and others on collaboration in shared-video systems.
Groups of participants performed tasks in a lab. This allowed direct observation of their use of the technology. I was present during each trial, and took notes and talked with participants. At the end of each trial I led participants in a group discussion about the experience, for which I prepared a list of questions.

To enable analysis after the trials I recorded participants’ screen output and voices. Using these I conducted conversation analysis, as well as quantitative analyses of deictic speech and use of Second Life’s virtual camera.

I checked emerging insights via participant observation and ethnographic methods such as attending building classes and competitions, discussing ideas with SL users discovered there and on the SL user forum.

**Study 4: Voice in a social world**

The aim of this study was to test the ideas that had emerged in earlier studies by broadening the range of virtual world scenarios in which data were gathered. Second Life was an ideal test system for this final case as it was a non-game VW which was used for diverse purposes. Furthermore its vendor had just added a voice channel, which had provoked controversy among users.

As in study 1, I interviewed in depth seven regular users of voice in the test system and conducting a focus group.

I monitored several out-of-world discussions - the official SL user forum, a popular education-in-SL mailing-list, and several prominent blogs - saving relevant threads for later analysis.

I became a regular user of Second Life and was engaged in several projects that necessitated dealing with problems of voice and text communication in my own work practice. These included membership of working groups which held regular meetings in SL, organization of a weekly in-world seminar, remote delivery of a lecture to an undergraduate class, and participation in mixed-reality music performances. This intensive period of participant research allowed me to develop insights and cross-check the experiences reported by informants.

The methods I used in my four studies are summarized in table 3.
3.6 Presentation of results

I have presented the results of each study in the manner suggested by Glaser and Strauss (1967) and other texts on grounded theory. Glaser and Strauss discussed the problem of 'conveying credibility’ during report writing, which requires that the researcher present his derived theoretical framework and support it with data. I have presented the substantive theory emerging from each case in the discussion section of each study chapter. Chapter 8 presents overall conclusions derived from all four studies (which can be considered substantive or formal theory in Glaser and Strauss’s terms). To support these conclusions I present excerpts of my data in the results section of each study chapter and in appendices D to G. In presenting data within the study chapters I have followed the advice of Glaser and Strauss who suggest that the author:

… describe the data of the social world studied so vividly that the reader, like the researchers, can almost literally see and hear its people – but always in relation to the theory. The standard approach to this problem is to present data as evidence for conclusions, thus indicating how the analyst obtained the theory for his data. Since qualitative data do not lend themselves to ready summary, however, the analyst usually presents characteristic illustrations. … He can quote directly from interviews or conversations that he has overheard. He can include dramatic segments of his on-the-spot field notes. He can quote telling phrase dropped by informants … (1967, p. 228)

Thus in each study chapter, the results section is organized into subsections corresponding to the themes that arose during coding. Each theme is illustrated with data (usually quotations from interviews), interspersed with text (written by me) which describes the concept that the

<table>
<thead>
<tr>
<th>Study</th>
<th>Interviews and focus-groups</th>
<th>Diaries</th>
<th>Observation</th>
<th>Video</th>
<th>CA</th>
<th>User forums</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>4</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

*Table 3: Data collection methods used in each study*
data illustrate. The discussion sections following each results section develop these concepts in detail and do not usually contain data themselves.

3.7 Conclusion

In chapter 3 I have described the methodology I adopted to answer my research question, the cases I chose, and the methods used in each case.

My approach was exploratory and informed by grounded theory and ethnography. The cases were selected to represent the wide diversity of virtual world systems, users and uses. I collected mostly qualitative, subjective data about user experience, but triangulated this with observation and quantitative measurement. I analysed data by coding it to discover themes.

The next four chapters describe the conduct and results of each case study in detail.
Chapter 4

Study 1: Voice in massively multiplayer online games

4.1 Background

The purpose of this study was to examine the influence of voice on the user experience of massively multiplayer online games.

Before this study it was known from the work of Gibbs et al. (2004) and Halloran et al. (2004) that voice introduced both benefits and problems for teams engaged in networked FPS games in which players negotiate virtual space in competition with other teams. As is the case with other games organized in space, such as sports, the relate location of team members matters in an FPS. To succeed requires coordinated movement of team members, which in turn requires efficient communication about time, space, plans and actions (Reeves et al., 2009). The benefits of voice included better coordination of groups, faster resolution of ambiguity, greater social presence of friends, and freeing up of hands for controlling other input devices (Gibbs et al., 2004; Halloran et al., 2004). Problems for FPS players included channel congestion in large groups, a greater opportunity for griefing and transmission of noise, inability to match voices with avatars, inability to know who might be listening to a transmission, and an unwillingness by shy people to use voice with strangers online. Speech, unlike text, could be received not just by those playing the game but by people physically co-located with users, such as family members or co-workers. Sound from the user’s surroundings, such as household noise, could be transmitted into the game, leading to unintended breaches of privacy (Gibbs et al., 2004).

The Xbox Live study of Gibbs et al. (2004) had shown that the benefits of voice depended on game type. In games lacking team play, such as driving simulations, voice was used only for trash-talking, which many users did not value. It appeared that the canonical scenario to which
voice was suited was coordinating team operations in virtual space. The further play diverged from this scenario, the less useful was voice.

### 4.1.1 Massively Multiplayer Online Games

While FPS games provide a virtual environment in which play occurs, it is not a persistent one. The environment contains only a handful of avatars, controlled by users who know each other and are often geographically co-located. By contrast, massively-multiplayer online games are persistent: for example, at the time of writing *Everquest* has just celebrated its 10th year of continuous existence. MMOGs typically host thousands or even millions of users simultaneously. Most of these of course do not know each other and are not physically co-located (Yee, 2006).

In most MMOGs users form long-term associations (usually called ‘guilds’) which are typically larger than FPS teams and often hierarchical in structure (Williams et al., 2006). Guild members collaborate in a number of game scenarios, the most notable of which is the group raid, a joint attack on an AI monster which offers significant rewards and cannot be performed by a lone user (Golub, 2010).

Because MMOG guilds may comprise hundreds of people, only a subset are likely to be logged in at a given time. Therefore some form of asynchronous communication is needed. (This might be conducted using communication tools outside of the virtual space.) Users also form temporary associations called ‘pickup groups’, which are usually smaller and may involve strangers not drawn from their guild.

The variety of activity is greater in MMOGs than in FPS games. While users engage in organized raids which are similar to the virtual combat of a FPS, they also spend time exploring the virtual world, socializing with team mates and with strangers passing by, and trading virtual possessions (Ducheneaut et al., 2006). While in both MMOGs and FPSs, users need to cooperate and communicate to achieve goals, the social interaction in MMOG worlds, populated with millions of users over an extended period of time, is richer and more complex. For many users the game is primarily a social experience – a ‘third place’ (Wadley et al., 2003; Steinkuehler and Williams, 2006; Ducheneaut et al., 2007) – which makes the ability to conduct convivial communication particularly important.

To summarize, communication in MMOGs occurs on both long and short time scales, over long and short virtual distances, synchronously and asynchronously, and between users both
known and unknown to each other. The prior work on mediated communication reviewed in chapter 2 indicated that some of these scenarios – for example asynchronous communication and communication with strangers – can be problematic for voice. Until 2005, MMOGs, like MUDs before them, had supported different communication scenarios by providing several different text channels - whisper, vicinity, party, guild etc - with different propagation rules, presented to the user simultaneously in separate text boxes on the screen. It was not clear how this might be achieved in a voice-enabled environment.

4.1.2 Voice is introduced to MMOGs

Voice communication was introduced to PC-based networked gaming through third-party add-on products such as Roger-Wilco. Microsoft’s Xbox Live game network, released in 2003, took a radical step in providing a voice channel but no text channel. It seemed to some at this time that users had voted for voice as their preferred method of communication in game environments. Richard Bartle sparked debate in 2003 by predicting that the burgeoning popularity of voice among videogame players might spread to MMOGs and have a detrimental effect on immersion and pseudonymity:

If you introduce reality into a virtual world, it’s no longer a virtual world: it’s just an adjunct to the real world. It ceases to be a place, and reverts to being a medium. (Bartle, 2003b)

Bartle’s prediction that voice would be widely adopted by MMOG users proved to be correct. By 2005, World of Warcraft had become the most popular MMOG in the US, partly by attracting a young audience that also played FPSs. Many of these brought existing team structures and practices, including third-party voice products, use of which became mandatory for membership in many high-achieving WoW guilds.

In 2005, Dungeons and Dragons Online (DDO) became the first MMOG to be released with voice communication facilities integrated into the game software. Several other voice-enabled MMOGs followed, such as Lord of the Rings Online (LOTRO) and EVE Online, all released in 2005/6. A built-in voice channel was added to World of Warcraft in 2007.

While many MMOG users embraced voice, some expressed concerns about privacy and pseudonymity. Designers, researchers and users on the TerraNova and Slashdot blogs debated the potential drawbacks of implementing voice in virtual worlds. Some felt that by making a user’s gender, age and ethnicity clear, real-world prejudice, harassment, domineering and
hierarchies, which were previously minimized by the pseudonymity of virtual worlds – and which some MMOG users sought to escape - might consolidate within virtual worlds. The separation between users and their characters might be narrowed, reducing the quality or even the possibility of role-play. People who embraced virtual worlds because a disability or abnormality stopped them fully engaging in the physical world might find this avenue for social participation cut off.

At the time of my study, voice was a recent arrival in MMOGs, and despite this controversy, no research had been published on how it was being used in this context. I perceived a need for a rich understanding of how voice influences the user experience in MMOGs.

4.1.3 Integrated voice in DDO

Using a third-party voice product that is not integrated with the VW is little different to taking part in a telephone conference-call. Users can log in and log out of the VW, their avatars can move, die, or be alone or in company, without affecting how they speak to each other. Conversely, the virtual world has no way of responding to activity in the voice channel. The channel must be separately paid for, and its configuration and operation is an extra burden above the group’s management of their virtual world use.

On the other hand, integrating the channel with the VW opens the possibility for new ways of using it. Channel management can be integrated with team management: for example membership lists needn’t be manually recreated. The game software can be made to visually indicate when a user spoke, or switched their voice on or off. The voice channel can take avatar location and orientation into consideration when propagating utterances between players. Users need not separately pay for nor install the voice channel.

The integrated voice channel in Dungeons and Dragons Online exploited some of these possibilities. When a user spoke, their name was highlighted on screen. When users formed a pickup group, a common voice channel was automatically assigned to them so that the team-mates could speak with each other in a “two way radio” format. Users couldn’t speak with non-team-mates, nor with ex-team-mates after the team had disbanded.

DDO’s thematic content was based on the pen-and-paper role-playing game Dungeons and Dragons, which in turn was loosely based on the Tolkien fictional universe which inspires many MMOGs. DDO’s game-play was similar to other MMOGs in that to progress in the game, users were required to battle monsters, some of which could only be defeated by
coordinated groups. DDO didn’t directly support guild management, but did support the formation of pickup-groups that lasted the duration of the current login session. (DDO indirectly supported long-term associations with a friends-list feature, and some users formed regular groups using coordination mechanisms external to the game.)

As with other MMOGs, play sessions typically lasted a few hours and (ideally) ended with the accomplishment of a game goal such as defeating a boss, and the disbursement of ‘loot’ to the group. DDO was unique in that loot was allocated automatically, preventing the problems some MMOGs had with arguments over post-raid disbursement.

4.2 Method

4.2.1 Aim

The aim of this study was to answer the research question:

RQ1: How does voice influence the user experience of massively multiplayer online games?

4.2.2 Participants

Three groups of five participants were asked to use voice in a massively multiplayer online game over a period of two months.

Participants were recruited through personal and professional contacts. About a third were known to me. Most were resident in Melbourne during the trial, however some resided in Brisbane and were interviewed there. The participants were aged between 15 and 40 and all but two were males: thus this sample was reasonably representative of MMOG players (Yee, 2006).

All participants were experienced MMOG users, and most had already used voice in online FPS games. This ensured that issues reported in this study were due to the interaction of voice with the MMOG genre or with the type of MMOG played, rather than the novelty of either MMOGs or voice itself.

One of the groups consisted of people who knew each other and had played games together prior to the trial. The other participants did not know each other prior to the trial. This enabled
a comparison between users who did and did not know each other offline, which prior research had indicated might be a key factor for voice.

Each participant was compensated $100 for their time and Internet use.

4.2.3 Materials

I used PC-based MMOGs as study platforms. I chose Dungeons and Dragons Online – described above – because it was at that time the only MMOG to have a voice channel built into the game software.

Participants playing DDO were supplied with a copy of DDO and a headset. They had to install the game onto their own PCs and get voice working, without help.

All participants were given a paper diary. This was a simple notebook without dates or times marked.

Interviews were recorded using a digital sound recorder. I videotaped the group interviews using a digital camera.

4.2.4 Approach

Two participant groups were assigned to play DDO. None of the group members had played this game before. The members of one DDO group knew each other prior to the study and had played other games together. The members of the second DDO group did not know each other and did not usually play together during the study, choosing instead to log in individually and join pickup groups.

The third group were already using third-party voice products in the popular MMOGs World of Warcraft and Everquest. They continued to use these during the study period and were interviewed to obtain a comparison across MMOGs and voice products.

The design of participant groups is illustrated in figure 7.
The participants played their assigned MMOG over a period of three months. They played whenever and as often as they wished, under their normal playing conditions in their own homes. Participants were physically separated; thus only mediated communication was possible.

Half way through the study, all participants were interviewed individually. At the end of the study they participated in focus-group interviews. All individual and group interviews were open and semi-structured and lasted one to two hours.

The research protocol was approved by the Department of Information Systems’ Human Ethics Advisory Group. All participants signed consent forms and were given Plain Language Statements. These are shown in appendix D. Interviews were recorded and preserved on audio cds, along with the participant diaries, in a locked cupboard to which only I have access.

**4.2.5 Data Collection**

The timeline in figure 8 illustrates the data-collection process for study 1.
Diaries

Participants were given a diary each and asked to spend 15 minutes to record events, ideas and feelings on each day of play. A sheet marked “How to use your diary” asked them to consider:

Was voice useful in communicating with other players? Was it better than typed text? How could the implementation of voice be improved? When and why did you use it? We are also interested in how you felt while using voice. Was it fun, awkward, embarrassing …? Please feel free to record both positive and negative impressions.

We would like your diary to be expressive of the game situations and activities in which you communicate with other players. You could include things like how and why you were communicating, what you were doing in the game at the time, and the purpose of the communicative exchange. We’re interested in the roles that other technologies (for example phone, SMS, email or instant messaging) played in your communications. We would also like to hear about why one method for communicating was chosen over another.

Please do not feel constrained in the mode you choose to express these situations and activities in your diary. Feel free to use whatever form or medium you feel comfortable with. We encourage you to use words, drawings or other forms as you see fit, to make a written and visual diary.

Individual interviews

Interviews were semi-structured and lasted about an hour. I sought to understand the participants’ use of voice and to collect their criticisms of existing implementations and suggestions for future ones. I used the following list of questions to stimulate discussion, but allowed conversation to flow according to what users found interesting.

- Did you like using voice-over-IP in the game?
- Was it better or worse than using text?
- How often did you use voice compared to text?
- Did your use of voice change during the three month study?
- Were their particular game situations in which either voice or text was preferred?
- How well was the voice communication implemented?
- How could voice be better implemented in MMOGs?
- Were any aspects of the user interface for voice frustrating?
- Do you prefer an ‘always-on’ or a ‘press-to-talk’ interface for voice?
- How much have you played MMOGs before?
- Have you used voice in games before? In what kinds of games? Was it useful?
• Is voice more useful in fast team games such as Counterstrike than in MMOGs?

• Is voice better for small or large groups of players?

• Is voice better for person-to-person, vicinity, quest, or entire-clan chat?

• Is voice better for some styles of play than others?

• Is voice better in some parts of game play than others?

• What communication technology would you like to see introduced in future MMOGs?

**Group interviews**

Participants took part in two focus-group sessions: these were also open-ended and lasted between one and two hours. One session was for group D1, and the other for groups D2 and M.

To prompt discussion I asked questions similar to those used in the individual interviews, and presented the following fictitious scenarios for participants to consider:

1. You walk into a busy tavern (30+ people). An in-character cleric casts a remove curse on you. You wish to thank her, as well as look for party members and meet some friends to discuss other quests. Only voice chat is available. Discuss how a future voice chat system could be implemented in such a situation. You will need to consider that there is the chatter of 30+ people around.

2. You played a mixed team of 'org mates' and some pick up players. One of the pick up players is acting quite rudely on voice and insulting your n00b carebear fighter as he calls it. Your mate is the team leader. You'd like to (a) calm down and (b) kick the bastard. How could you deal with this on voice?

3. You've been involved in a chaotic fight, focussing on staying alive and not hearing everything that was going on between your team mates. At the same time an out-of-party friend (guild-mate) sent you a Tell. You are aware that you were contacted but are unsure of the content of the communication. Now that you have some breather space, how would you deal with 'past voice chat' that you may have missed. Is it important to follow up? How does this compare with scrollable text logs?

4. Explore the multi-channel (raid, guild, tells, party, auction channel) capabilities of text, and contrast it with what voice offers. And can voice ever work for multiple channels? If yes, how so?

5. Could your current face-to-face role-playing sessions in other games ever be performed in a virtual world? Or would this be completely undesirable?

Samples of the data gathered in study 1 can be inspected in appendix D.
4.2.6 Data Analysis

I transcribed the interviews, coded the data into themes, and arranged these into a hierarchy. I checked my coding through discussion with co-researchers. My first analysis was used for a conference paper arising from this study (Wadley et al, 2007). The hierarchy of themes used for that paper can be seen in appendix D, figure 26.

Later while writing this thesis I recoded the data afresh, this time using NVivo. The passage of time and sensitivity to issues arising in later studies led me to refine the codes and I ended up with a larger number of themes.

I sought high level concepts within which to categorize these 18 themes. The most striking feature throughout the data was the depth of feeling participants held both for and against voice. Therefore I organized my data around these emotional responses. I created top-level categories to include feelings pro, con and ambivalent about voice. The fourth top-level category ‘Using voice’ contained general observations on how voice was used. Finally there was a category ‘Suggestions’ to contain improvements that participants wanted to see in future voice implementations. My final coding hierarchy can be seen in appendix D, figure 26.

These results are presented in the following section. Each result is illustrated with fragments of data, following the style suggested by Glaser and Strauss (1967). (See section 3.6 for discussion of presentation style.)

4.3 Findings

Study 1 yielded 18 themes which I organized into 4 top-level categories, representing participant feelings for, against, and ambivalent about voice, and a final category ‘using voice’.

In this section I present the 18 themes, grouped by category, with illustrative data for each. The higher-level themes are discussed in greater depth in section 4.4.

I begin with data on participant feelings in favour of voice.
4.3.1 The benefits of voice

Nearly all of the participants approved of communicating by voice – under the right conditions. After technical problems were resolved, participants found speaking to be easier, more natural, and more relaxing than typing. They felt that voice enhanced the user experience of MMOGs and added a new social dimension to the game that text communication could not deliver.

It was great fun. It made the game more intimate. So you’re not just playing a computer game. It’s more involved.

I jumped into WoW [where only text was available] for something to do and we thought this feels weird. It doesn’t feel right. We need voice. Without the voice there was just that difference in level.

Using text chat was ok – it just seemed a bit lacking. I wouldn’t have realized before using voice just how much of a difference it makes to playing the game.

It definitely adds to it.

Voice made communicating with other players easier. Conversations flowed more freely than in text, and more was communicated in a shorter space of time. Participants said they would be reluctant to revert to text and felt that without voice the game seemed awkward or unresponsive.

It just seems a lot more natural being able to talk to people and say “come over here, check this out”, as opposed to having to type stuff..

It’s really useful when you’re in a group. Finding your way around, asking where to go.

Participants emphasized that it was interaction with other people that drew them to online virtual worlds: “Yes, you wouldn’t play otherwise.” For some, voice enhanced this social interaction: “That’s what makes it. It’s not just functional, it’s social.” Some participants, especially those playing regularly with friends known to them offline, approved the greater social presence conveyed by voice, saying it made the game “feel like a living breathing party”. Voice made it feel as though fellow players were physically co-present: one participant comparing it to the feel of a LAN party.

In DDO we have a consistent group, six of us, they come together, most of the characters play together and they fill a niche. It’s almost like we’re in the same room.

Socially it adds the ability to talk to people, and it’s like they’re in the room there with you. There are times when we have what we call “LAN nights” when we get together at a person’s house, and sometimes playing DDO actually feels like that, where there’s a social aspect to it as well as the ability to play the game a lot better.
I think it’s also more social, because you’re playing but you’re also interacting as well.

Some participants felt that the use of voice helped bind members of a guild together, by making them feel as though they knew each other. In fact, players who didn’t use voice became ‘invisible’ to the group.

If I’d never talked to them on Vent, I probably would not be in the guild any more. I’d go to other guilds, just find a guild that I like, keep moving around to benefit myself. But because I talk to them on Vent I sort of know them, they help me and we help each other, and that’s why I stay in the guild.

If you see in the guild, people that don’t talk on voice, that just sit in the background, I don’t know their name, I never quest with them, they just sit in the background and get ignored, nobody ever notices them.

While voice certainly had disadvantages, most participants seemed to feel that on balance the advantages won out.

I’d rather use the voice than not use the voice. In parts it does ruin it. If they’re female, I’m like “woah you’re not female”, it’s a bit of a shock sometimes. But other than that I reckon it adds a lot to the game. Even when you’re by yourself and doing your own thing, just having somebody else to talk to, and ask questions. It’s like an online database, if there’s a quest, they’ve done it before, so you say “can you give me a hand with this?”, and they say “yeah you gotta do this and this …”.

For these voice users, the enhanced social atmosphere and the ability to ask for help with the game overrode concerns about diminished ability to role-play.

4.3.2 Voice is most useful during fast-paced action

All participants agreed that voice was superior to text during raids, during which coordination of fast-paced action was required.

Especially in quest scenarios, if there’s something that happens quickly, it’s a hell of a lot better to say ‘heal so-and-so’, or ‘there’s a trap ahead’, as opposed to trying to type out ‘trap’.

It adds the dimension of being able to play at a faster pace. You don’t need to type in “help” when you can just yell it. We seem to be able to coordinate our attacks a lot better.

Because speaking freed a user’s hands from typing, participants found that they could communicate with team-mates while carrying out other game actions such as moving and fighting. Participants reported incidents in which this made the difference between success and failure in the game.

I do use text chat a bit, but it’s impossible when you’re in a fast-flowing game, you’ve got guys running at you and you’re having to heal people and cast spells, it’s impossible to chat at the same time. It’s a lot easier to just talk, rather than stop to chat then start playing again.
If you have to stop and type you’re probably not going to be responding to things. When things are happening you don’t have time to take your hands off the keyboard.

When you’re hitting something you can’t say “come help me”, because if you stop to type that you’re dead. It’s better for a game that’s real time.

As a result of their improved ability to coordinate, adopting voice improved participants’ team performances and led to greater success in the MMOG.

A lot of times in WoW when we fell down it was because we couldn’t communicate that information fast enough, whereas in DDO, in some of the big battles, I expected us to go down quite quickly but I found that we were actually doing them quite easily. I couldn’t figure out why until it occurred to me, the ability to communicate quickly was one of the key factors in some of those groups.

In WoW we’re using typed text for most communication, but when we’re doing high-level instances, because of battle orders, then we’re using voice.

Yes I use Ventrilo. As a matter of fact it’s essential for a successful run to have voice communication.

Coordination was especially important among large teams. Voice was a better way for raid leaders to issue directions to subordinates, and for wounded users to call for help. It was also well suited to the discussion and planning that took place prior to raids, and to negotiating the distribution of loot afterwards.

When we raid we’re raiding with at least 40 players, so there has to be a degree of command and control, or some of these bosses are really hard, they just don’t go down.

Voice enabled new kinds of coordination and tactics for raiding teams. Most importantly it enabled real-time coordination among player classes, and allowed plans to be changed in an ad-hoc manner as situations arose.

I really love the speed at which voice communication can relay battle commands. So if someone is on the verge of dying you can say “character A needs healing” and so the cleric can heal that character before they die. Whereas by the time you could type it in, that character could be dead. That was fantastic.

I think it allows certain tactical things too that we couldn’t do otherwise.

We change our tactics on the fly and coordinate that by voice, which would just be impossible via text.

One participant reported that voice was mandatory in the large, high-performance raiding guilds in WoW, where it had “an obvious advantage”.

All the top guilds use voice communication. It’s just so much easier.

On a raid, they won’t let you into the raid unless you’re on Ventrilo. They say “use it or don’t come”.

Not everyone in the guild is using voice, but a very large majority. It’s caught on quite a bit.
The officers want people to talk on Vent, and they’re a lot more inclined to help if you do.

Another participant liked using voice in DDO simply because she disliked text communication in general.

I don’t even like SMSing. I’m not a good example of a texting person. I don’t like putting things down in text.

Many MMOG players are highly motivated to succeed during raids, and for these people voice presented clear advantages. This concords with Golub’s (2010) findings on the use of voice by high-achieving raiding guilds.

### 4.3.3 Easy sociability

Many participants appreciated not only their enhanced ability to coordinate battles but the ease with which they could enact other forms of interaction with fellow players. This extended to “off topic” socializing unrelated do the game.

Voice is useful in proportion to the need to talk to other players in the game. When you’re looking for a group, or trying to organize where everyone should meet up, or you’re trying to strategize. Anytime when everyone knows what to do already, it’s fairly redundant.

We use it for just chatting about stuff we’ve done in the daytime, and having a joke, that sort of stuff.

Voice made it easier to make off-the-cuff comments: therefore these were more likely to be made. Many participants felt that a valuable social layer was added to their gameplay.

While you’re in the dungeon crawl or whatever, the comments … That’s what makes it as far as I’m concerned. That’s the thing that I like. It might happen [in text] but it’s scrolled by before I can see it, and there’s no way I could make those comments if I’m not going to able to type while I’m playing.

I reckon when it’s in text, the comments come before and after the big fight, then while you’re actually fighting it’s all quiet because everyone’s doing stuff. But in voice you have the chance to make comments off the cuff.

However not all participants appreciated off-topic talk. The easy sociability of a voice channel could lead to irrelevant or distracting conversation. A chatty user could easily dominate the channel and annoy team-mates with what one participant described as “verbal diarrhoea”.

What was really irritating was that sometimes the small talk wasn’t really game-related. It was just kind of people messing around, trying to sound self-important or trying to be cool, and it diminished to the point where I would just turn the volume off.

There was this one guy who let us know exactly everything he did. Go through here, go open that door, everything he did was told to us.
I remember this one guy who kept talking about “Oh, I’ve been playing this game for the past year, and I’ve got 5 level-10 characters and I’ve deleted four of them and I’ve started them off again because … blah blah blah”. Dude just shut the hell up because I don’t care …

In other words, there were two sides to the easy sociability that voice communication brought to online gaming. In the next few subsections I present findings concerned with problems that participants encountered in their use of voice.

### 4.3.4 Immediacy, emotional impact, and abuse

Despite their broad acclaim for the benefits of voice, participants were equivocal about some of the phenomena that voice engendered. For example voice’s immediacy and the extra presence it conveyed increased the emotional impact of messages. On the one hand, this could be a force for good in team management.

If I was disciplining someone, if someone was being rude, I’d love to be able to look at them, and have them look at me, and say “listen, if this keeps up you’re out of the group”. I’d love to do that face-to-face. I’d love to be able to just get that seriousness across.

If someone types it you might just ignore it. But if they say it you react to it.

On the other hand, the immediacy and emotional impact of voice could lead to the transmission of more criticism and negative emotion. Voice made criticism feel more real to the receiver. Flame wars could be started and sustained more easily, and this could damage team cohesion.

You’re sitting there running a raid and something happens, and from chat you get ‘WTF’. In the verbal chat, when someone actually says “What the f***!”, and says it with exclamation, it gives a whole new emphasis to the experience.

Negative side I found is that if someone is irritated it comes out in the ways that they speak to the rest of the group, so that can aggravate people and cause dissension in the group.

If you’ve got voice you’re going to vocalize, but if you’ve got text and you’re upset, you’ve got that extra level of barrier.

There’s certainly something different about shouting “shut up you dickhead” into your microphone and typing it.

Participants noted that while text users were more likely to think for a moment before ‘pressing send’, voice users tended to speak without first considering the implications.
The difference between voice and chat is that with chat people have to think a bit more about what they’re typing, whereas with voice you say straight away, there’s no thinking, it just spews out. I don’t know if you’d call it a disadvantage. But you can type something and think “no that’s a bit harsh” [before you press Send]. You can cool down.

On the other hand some felt that instant communication could heal rifts. A team member who made a mistake could apologize immediately, and their sincerity would be perceived.

I can remember screwing something up in WoW … you’re too busy running for your life to type “sorry guys, bad call”, whereas with voice it’s like “screwed up, watch out, incoming”.

I think in that case, I’ve done this, “oh shit!”. You’ve verbalized that you’ve screwed up before they even know.

A few participants felt that, in fact, people were more polite when using voice than using text.

People treat voice like a phone conversation – they have phone etiquette.

I haven’t noticed people being rude at all.

Nobody’s tried to chat me up. Nobody’s made any sexist comments. I had one, but it was a joke.

One participant believed that whether swearing would cause offence depended on whether the users of the channel were close friends.

As an adult sometimes you’re going to be a bit crude or say a crude joke or a sexual reference or something nasty. It’s put in context, as an adult, if you’re with someone you know, with your guild, you say “you wiped us you dumb shit” and they’re like “f**** off you bastards”. In real life, if you say something in a light manner and they respond in a light manner, there’s no offence because you’re friends. If I’m walking down the street and I bump into someone and they say ‘watch it dumbass” I’m going to be tempted to knock his block off.

Offensive language was therefore likely to be problematic primarily when encountering people one didn’t know online.

In its most extreme form, impolite use of the voice channel deteriorated to abuse, flaming or griefing. Here, the emotional power of voice was a problem.

I play WoW a lot, and we use the text chat system in WoW, and that can get nasty. But if someone is irritated, you don’t get that initial sense that they’re irritated. They’re just words on a screen, as opposed to actually hearing the human voice.

If you can just walk into a tavern and yell things out, do you see how many idiots out there would abuse that? Some little kid would just start ranting off swear-words or racial comments.

I did something I wasn’t supposed to do. The raid leader got really irritated with me and basically voiced his irritation. I wanted to tell the guy to go jump in the lake. Instead I typed in saying “oops, I’m having microphone problems and it looks like I’m going to have to relog”, and I basically left the game and left the group and that was it.
Participants noted that recordings of MMOG users losing their temper at team-mates had become commonplace on the Internet, offering the well-known ‘Leeroy’ recording from WoW as an example. However it was felt that flaming was more of a problem for pickup groups than for people who played together regularly.

I think it depends on the makeup of the group. I’ve never had an issue when it’s people I know playing and the minority have been pickup people. A couple of times we’ve had people we’ve picked up in DDO and they’ve been fine. But if I’ve joined and I’ve been the picked up person, it’s been completely different – I’ve seen some really atrocious misbehaviour.

Some felt that sexual and racial harassment, already prevalent in ‘public’ (unconstrained) text channels within MMOGs, could become worse in voice channels.

I know in DDO some girls said they wouldn’t use voice because they’d get harassed. So on the flip side, because of that anonymity, it seems to bring out the worst in some people. So if there was the ability to walk up and click on people, I suspect you’d get that aspect coming out. It would start a lot more harassment.

Blizzard changed their looking-for-group channel [in WoW] from area-specific to world-wide. This has been changed for the past 2-3 months. … So you have this madness, idiocy, of people being rude, racial comments … They’ve got a bigger audience … If it was like DDO and you’ve got integrated voice-chat, and they tried to do the same thing to have this world-channel voice chat, it would be totally crazy. It just wouldn’t work.

You’re not going into a game to be persecuted because you’re a Jew or you’re gay, you’re going there to have fun.

You do something like that in real life you might go to court, you might go to jail or get fined money. But in MMOs you might get banned for a day or something. So you just create a new character, or buy a new account, so the penalties don’t reflect what they are in real life, so people feel they can get away with it.

These comments indicate that the problem of abusive language in large virtual worlds was not restricted merely to the utterance of swear words, but extended to race, religions and gender-based vilification. Participants were concerned that the use of voice could make this situation worse.

Some participants expressed concern that children might be exposed to abusive or sexual language. Whereas it is easy to automatically filter text using software, this would be harder if not impossible in voice.

There might be an 11 year old girl online. Now if there’s a 45 year old biker on there swearing with his mates on voice channel, that would be an education for them. She’d freak out and get lost.

---

8 This can be viewed at http://www.youtube.com/watch?v=LkCNJRfSZBU
I know WoW has a profanity filter in the text you can activate, so you can filter out the swearing. In voice you don’t have that luxury.

Despite this concern, some participants resisted the idea of censoring the voice channel, and felt that preventing exposure to profanity was the responsibility of parents.

I’d actually be pissed off if someone was censoring what was said by removing profanity. Not that I think every word I say is golden. But just that, if they can do that to what I’m saying, what am I not hearing that everyone else is saying? And to me in some way profanity is part of the message that you’re communicating. It’s a choice that you make to communicate a particular emotion.

It doesn’t matter if it’s a computer game, or a TV show, or the Internet or whatever, in regards to us having to be careful of what we’re saying because there might be a 10-year-old on, if the game’s M-rated or PG-rated or whatever, it really is the parents’ responsibility to be policing that.

Some participants encountered players who were quite creative in their (ab)use of the channel.

One encountered a team-mate who sang.

I had one guy singing. Singing a song that he made up in context. Role-playing. He was singing in a pub. We’d been in a battle and it was one of those early ones where people just ran around and did their own thing, ignoring everyone else, and we all died as a consequence. We were resurrected at a tavern and he started singing a really really long song while we were healing, about how we just wanted to heal, any kind of heal would have done. It was ok for a while but he just kept going, and having more and more of a dig at the cleric.

Abuse of the DDO voice channel even extended to a player attempting to convert team-mates to their religion.

There’s this one guy who started talking religion, and that was not good. Somebody else said “I don’t believe in god”, and he said “that’s ok, god believes in you”. It was just yuk, and it ruined it because we had a really good game that time. That guy who was saying that stuff had been really good in the game. Maybe he could get away with it if he’d been a cleric and in context but he wasn’t. [laughs] It spoilt the game a bit for me.

Thus the immediacy and ease of using voice presented both positives and negatives for MMOG players. Voice provided an immediate, intuitive social connection between geographically-separated users of large virtual worlds, and this dramatically enhanced the sociability of experience and users’ effectiveness in competition. These benefits accrued especially to users who already knew each other offline and were operating as a team in fast-paced joint actions. However the benefits could be outweighed by a significant set of disadvantages when users were not familiar with each other, culminating in problems of griefing and deliberate harassment.

A related set of problems stemmed from the ability of voice to convey information, not just from but about users - a key finding of media richness research.
4.3.5 Personal characteristics exposed

Another property of voice about which participants were equivocal was its capacity to convey the personal characteristics of the people speaking. Several participants noted that they could discern facts about users by listening to their voice that were less obvious when communicating in text.

One participant (an American resident in Melbourne) found that when his WoW guild adopted TeamSpeak, he quickly learned new facts about his team-mates.

The same woman, we were on TS, this was the first time I’d heard her voice. She was obviously a middle-aged woman from the South, she had a very distinct Southern accent, but she had a really deep crackly voice. My first thought was ‘smoker’ because my mom has a deep crackly voice. When I heard her voice for the first time, the image I had of her from her personality were two distinct things. It was her voice and her attitude, because she was actually cussing quite a lot. My original conception was a middle class woman - now to a lower class. Not to make a judgment, but more understanding who she is on the other side of the microphone.

It’s actually quite interesting - your preconceptions of what people are like - because hearing them tells you a lot more about them. You can tell their gender and their age from their voice. Even nationality

In virtual worlds where team membership often crossed national boundaries, voice exposed cultural backgrounds that were usually obscured by text.

One time I said “bloody spear” because my character got speared, and someone said “oh you must be Irish or Australian”, and so we all had this chat about where we’re all from.

We’ve got a guy from the Philippines – he’s got a bit of an accent. And we’ve got guys from Canada, America, lots of Australian guys.

I haven’t had too many problems with it, the main thing is just the comments - “Oh you’re an Aussie are you?” – the majority of people seem to be American on this server.

And whereas communication by text usually hid a user’s age, users adopting voice sometimes realized that they had been playing the MMOG with children. Sometimes the effect was merely comical, as when the voices of young children mismatched the appearance of their avatars.

You can tell kids. They’ve got little squeaky voices. … There was this huge barbarian tough-guy, massively strong, really low in intelligence, and he speaks with quite a high pitched voice through his little mouth.

However one participant, a school teacher, pointed out that exposing the age of children in online games could be dangerous.
If you’ve got a kid playing a game with adults, and you’ve got a pedophile on the other end thinking “oh there’s a kid on here, I know that they’re a kid because of the sound of the voice, I can take advantage of that and try to get into this kid’s confidence, and go meet this kid.”

Exposure of personal characteristics had other ramifications for MMOG use, such as when users engaged in role- or gender-play – these are discussed below.

### 4.3.6 Speaking with strangers

A number of problems arose because of the unease that some voice users experienced in situations where they might be overheard by people they didn’t know. It appeared that this was much more of a problem for voice than for text.

Some participants found speaking with strangers to be uncomfortable.

I don’t know the people outside Everquest, in real life. At first I found [voice] a bit intimidating. I’d go on and be very reluctant to say anything.

Some people don’t like it – they feel uncomfortable with talking to people. They’re a bit shy.

A viewpoint frequently offered was that voice was better for friends who collaborated over long periods than for pickup groups who are usually meeting for the first time.

For new people joining the guild, they just listen a bit and try and work out who’s talking. We’ve been playing for two months now so we know each other’s names. I know their wives’ names and girlfriends’ names, what school they go to. There are two couples in America, they both play at the same time, they’re husband and wife and they go around helping each other.

I think it’s more about your personality and your familiarity with the other people. If you know them well you’ll just chatter away.

I think with us, because we’ve known each other for a fair while, it makes a difference with regards to funny things, because we’re all familiar with each other and so we joke around anyway. And the flip side, the agro side not the humour side, that never seems to come up. That familiarity makes a huge difference I think.

It’s good for DDO because we’ve got a consistent group.

In a sense it reflects real life in that you wouldn’t walk to someone in the street and just ‘blah’. But if you are waiting in a queue - like I find that at the Melbourne International Film Festival, because of the environment of it, you’re waiting in a queue to see the next film, if you notice someone having seen a different film, you might say “oh what did you think of that film?”

I’m not entirely happy with [voice]. There’s a lot of things I don’t like about it. If I’m playing with my close friends it makes it a lot easier.

Some participants encountered users who didn’t want to use voice in pickup-groups because they were uncomfortable speaking with strangers.
Especially in the early days we found only one or two people in a pickup group using voice. It didn’t seem to get used that heavily. It seemed to be mainly [name] and I who were doing all the talking. I don’t know if that was because we knew each other. We were just having general fun banter, role-playing, and doing the dungeons. You might occasionally get something but usually we had to direct a question to a person to get a response out of them. I don’t know if they just weren’t relaxed to talk.

In fact some participants said that using voice in their teams led to a reluctance within the team to include strangers in their game play.

I think from our point of view, voice has been restrictive, because one of the guys doesn’t like partying with people he doesn’t know. Occasionally someone wants to join, and he says “I’d prefer they don’t”.

The downside was it made me less inclined to talk to strangers.

This depended on personality, with shy users in particular being reluctant to use voice when teaming with people they didn’t know.

I’m the type of person who doesn’t like initial contacts. I’m a bit on the shy side. I don’t like to go to a bar, walk up to someone and start talking to them. So to use written chat makes me a lot more comfortable with someone I don’t know. The first time I used DDO I had some difficulty getting used to using the headset and microphone in a PUG group, with people I didn’t know.

One participant felt that while usually everyone in a group would be willing to speak about game activities, only people who knew each other well would speak about issues and events in their offline lives.

A lot of our guild members live in the same suburb, same town, and so they kind of see each other, in person, whereas someone like me feels … because you don’t know these people, don’t know what they look like, they become … I don’t want to use the word insignificant, but it’s a different feeling that you get.

One participant felt that activities like gender-play (using an opposite-gender avatar) worked better with people he knew.

When you create characters you’ve already got that consent of, you already know this group’s all right with me rolling a female character, they’re not going to think that I must be queer because I want to play a female character, because you tend to know the people. Whereas if you’re online and playing a female character, you don’t know how people are going to react to it.

This and a number of other data suggested that some MMOG players considered gender-play to carry a stigma, and that it was this that led them to conduct these practices only among people they knew and trusted.

Reluctance to voice with strangers depended on the game activity currently underway. It appeared that during fast-paced teamwork, the utility of voice could outweigh any discomfort attached to voice use.

If you run across somebody and they ask you for a buff, you don’t want to hear some big growly voice coming from this little avatar. However, once you’re in a team with some people, you’ve got a
motive between you, you want to kill the alien, you need to take down a target, this is something that needs coordination, you can work together and get past that.

So for general one-on-one communication with strangers, text might be preferred, but in a raid the utility of voice was too attractive to pass up.

Some participants suggested that reluctance to speak with strangers would be an important issue if an MMOG had “vicinity” voice (such as the system examined in the next chapter). Their discomfort at being overheard by strangers meant that when using a spatial voice system, they would want to be notified when another user’s avatar had moved within hearing range of their conversation.

A DDO participant found that most of the people he grouped with were listening to the voice channel (he could tell because an icon on the screen showed who had voice enabled), but fewer were prepared to speak. This participant took cues from team-mates: if others spoke, he did too, otherwise he did not. He felt that when users didn’t know each other, their conversation was shallower and their speech less frequent, so that the voice channel was used merely for giving commands and directions. However there was greater anxiety about “lurking” (listening without transmitting) in voice than in text. This felt more like eavesdropping - something that one shouldn’t do - whereas in text it seemed normal to scan the conversation flowing past without interrupting it.

One participant in a regular guild said his favourite use of voice was to chat off-topic with his friends while they explored the virtual world, or during boring, repetitive play such as trade-skillling and grinding. Another participant who usually played in pickup groups reported “occasional off-topic banter at the end of raids”, but “didn’t encounter anything particularly meaningful”.

When a MMOG player is using voice, eavesdropping can happen not only within the virtual space but in the player’s physical context. One participant felt uncomfortable speaking with strangers when family members were with him at home – that is, not because the strangers could overhear but because his family could. This anxiety did not apply to text communication.

Usually it’s me sitting in the corner chatting, and my wife over there reading a book or watching television a few meters away, and I would feel weird sitting there talking to two or three people I didn’t know. So it definitely made me less likely to join a pickup group. In WoW [using text] there’d be lots of people who you didn’t really know, but you let them in.

On the other hand, some DDO users, who were unable to speak with people outside their teams, wished for the ability to speak with “random strangers” whom they met while travelling
in the virtual world. For example, they wanted to use voice to ask for healing, and then thank obliging strangers, in a friendly, personal way.

One participant felt that speaking with strangers usually worked well enough.

I think it depends on the makeup of the group. I’ve never had an issue when it’s people I know playing and the minority have been pickup people. A couple of times we’ve had people we’ve picked up in DDO and they’ve been fine.

It appears then that anxiety over using voice with interlocutors unknown is widespread but not universal, and probably depends on personality factors such as shyness or introversion.

In the next few subsections I present findings about the user experience of communicating by voice in MMOGs.

4.3.7 Role-play while using voice

Participants debated whether role-play could be conducted when communicating by voice. The main problem perceived was that voice would make their personal characteristics more discernible, ruining the illusion required for role-play. There seemed to be a widespread understanding among MMOG players that role-play required the relative leanness of a text medium.

In WoW I was on a role-playing server for a while, and we weren’t using voice at the time, but the suggestion was to start using voice for some of the instances. Another person and I said we weren’t sure if we wanted to use voice, because we tend to role-play. We said it will dispel the illusion, because you’ll hear me with an Australian accent and a male voice, and this other person was also playing a female character.

We haven’t imported any role-play elements into our voice.

A couple of guys would role-play, but no-one would role-play verbally. It didn’t happen often. They pretty much relied on text.

And there was another guy who spoke a lot too, but he was role-playing, so he didn’t actually vocalize anything, he just typed reams.

Another participant was happy to role-play in text but felt it would be embarrassing to attempt it with voice, unless everyone in the group was speaking in character.

One participant who described himself as “an occasional role-player” said he used text because his collaborators spoke off-topic too often if they used voice. This is again a case of dispelling the illusion, but via a different mechanism – off-topic chat rather than exposure of identity.
Many participants seemed to define role-play according to whether users were feigning the voice of a game character, and by this definition claimed to have encountered few genuine role-players in DDO or other MMOGs.

I’ve never spoken to anyone who’s tried to speak in character using their voice. In text even it’s very rare – I can’t imagine somebody trying to pull it off with their voice.

People seem to talk in their own voice – they’re not trying to role-play and they’re not worried about the fact that it may be breaking the illusion.

There’s probably one guy I’ve encountered who was role-playing truly, and he didn’t use voice at all. He was role-playing with text. He was a dwarf, axe-wielding, and going “och aye” and all that. He had voice on, but he never used it.

However role-play was a controversial topic, with participants offering varying interpretations of what it meant and how popular the activity actually was in these supposedly ‘role-playing’ games. Few participants admitted to engaging in role-play or even witnessing it. Again I got the impression that MMOG users perceived something of a stigma attached to role-playing.

Most of the guys I talk to aren’t that into the role-play – they’re more about playing the game, getting the good items, playing it as a game.

I find with the guys, we do role play a little bit, we joke around about the characters but, I don’t know if I’d call it full on role playing.

I’ve never done that [role play] at all. I just don’t like the idea of meeting up with strangers and …

I’ve seen some people go around talking in Old English and they really pretend that they are the characters. I wouldn’t get into that but I like looking at the scenery and that sort of stuff.

I guess the role-play at my age only comes in after the second glass of wine, and then it’s only an accent and how wonderful my character is. In voice we’ve been talking so much about what happened during the day, or talking so much, that to get into character would be a bit of a stretch.

However some participants argued that you didn’t need to feign an accent or use ‘Tolkien-speak’ in order to role-play, claiming that this was a common misunderstanding among MMOG players. They preferred to define role-play in terms of undertaking an assigned role in the virtual world.

I think people misconstrue role-playing a lot. They think role-playing means saying ‘aargg’ a lot. [laughter] Even if it is about how you act, people think you have to speak that way as well, to role-play. Whereas you can actually just speak normally, and be role-playing in the way that you act. Some people might talk normally, but if they’re an evil character they might not stop to help the guy on the side of the road, or if they’re a good character they might. It’s that kind of decision that is role-playing because they’re doing what their character would do rather than what they would do.

You’re not acting, you’re not putting on a voice. I haven’t heard anyone who’s putting on a voice. Not pretending to be anything that they’re not, but responding to what’s happening in the world. The online world.
Another thought of role-play in terms of depth of immersion in the virtual world and the action taking place in it.

It’s interesting when you say “playing in character”, ‘cause I like to be in character, not in a heavy duty way, but as in pretending it’s me that’s walking around and doing things. The things I say are really part of that. “Oh wow look at that!” and so on, but not really putting on a voice. And I prefer it a lot when people I’m playing with are doing that level of role-play, so that when something happens they respond to it, as if it’s something that’s just happened, rather than just having a chat.

This participant felt that there were some game situations where voice seemed inappropriate and reduced immersion in the virtual world; for example engaging in light banter during moments of drama and death.

By far the most common form of role-play discussed by the participants was using an avatar of opposite gender. Confirming Yee (2005) and Blinka (2008), this usually meant males playing females. I detected a curious contradiction around this phenomenon. Men who played female avatars did not perceive this as problematic. Most discussed it in a matter-of-fact manner. The reasons they gave for this practice referred to aesthetics and game mechanics rather than to the exploration of alternative gender roles, confirming the reports of MacCallum-Stewart (2008).

However some described their encounters with other users’ gender-play as weird or uncomfortable.

It’s so weird when you’re playing in a group and there’s an obviously female character with a male voice. I tried to play a female character once and I couldn’t do it because I felt like I was cross-dressing. It felt wrong. I’m not a homophobe and I’m not gay either. It was just really uncomfortable.

Some simply found that gender-play made the job of matching the voices that one heard with the avatars that one saw more difficult.

I get completely confused when there’s a female-looking character with a male voice. It’s like “you’re a male, just don’t …”. I haven’t asked – I might in future though [laughs].

One female participant said she used a female avatar specifically so that her voice and appearance would match.

Participants who engaged in gender-play spoke of problems that arose when they adopted voice and exposed their offline gender to others.

---

9 I am using the term ‘gender-play’ to describe use of an opposite-gender avatar. I am wary that the term implies that the user is role-playing a person of the opposite gender, when my data and previous studies such as MacCallum-Stewart (2008) suggests that the practice is merely aesthetic or utilitarian. Yee’s “gender-bending” probably suffers from this problem also. But I have not found a better term that is still short.
We got startled reactions from the rest of the party. They were people we’d been venturing with for 30 or 40 levels and didn’t realize we weren’t female. Two of the other females were female, and they were quite surprised that we weren’t. They had no issue with the fact that we were male, but it destroyed the illusion, and that’s why we were against the voice. We could roleplay as females but the difference between typing it in and people imaging it in their heads, as opposed to having to hear the voice and then pretend that yes, that’s a female you were talking with.

If you’re sitting around just having a bit of a chat with somebody, you might not want to be talking with them using real voices. You might want to go with the role-playing idea that they are a male when they’re actually a female or vice-versa.

This discomfort seemed to depend on whether the gender-player knew the people they were engaging with, which suggests some degree of anxiety about being ‘found out’: “Myself and a couple of the other guys were playing female characters, and we thought that it would detract from the immersion.”

Gender is not the only way in which MMOG users’ online and offline appearances can diverge. Some commented that many MMOG characters are large, brutal warriors: a characteristic that players typically did not share in real life, and as for gender, the illusion could be broken by voice. One user commented: “My character looks scary, but my voice isn’t – how’s that going to work?”

Not everyone agreed that voice was damaging for role-play. Some of the male participants had participated previously in pen-and-paper (table-top) role-playing games such as Dungeons and Dragons. In the face-to-face role-playing scenario, hearing a fellow player’s voice was routine, and playing a character of a different gender or species worked perfectly well.

The game I’m playing on the weekends I’m playing a French character. I slip in and out of a French accent depending on the situation. If you’re playing sci-fi games you might be playing an alien race.

Pretty much all the guys I role play with once a week in pen-and-paper games. It’s a different level when we play DDO. It’s one step back. It’s not that we’re not role playing at all, but it’s not at the same level as when we pen-and-paper role play.

Participants discussed why role-playing in voice was possible “face to face” but not online. Again it seemed to come down to the concern that fellow players on the Internet were unknown, and possibly even eavesdropping.

When you play characters, you’ve already got that consent. There’ll be no hassles if I roll a female character. They’re not going to think I’m queer because I want to play a female character, because you tend to know the people. Whereas when you’re online and playing a female character, you don’t know how people are going to react.

Some participants were not concerned about the impact voice might have on role-playing in MMOGs.
I’m not much of a role-player, but you do notice that aspect of male players with female characters. But then people seem to talk in their own voice – they’re not trying to role play and they’re not worried about the fact that it may be breaking the illusion.

Several participants were enthusiastic about the idea of voice-changing software that would allow them to sound like the character they had chosen to play. However none had encountered such a system.

I think game designers should focus on technologies that alters people’s voices, especially with role-play. I tend to personify all my characters. It’s part of making the game fun. If I’m playing a dwarf I’d love to sound like a dwarf. If I’m playing an elf I’d love to sound like an elf. And when characters talk back to me I’d like to hear the same accent or dialect. That would be a lot more immersive than just using your average human voice. I’m sure there could be some sort of conversion so a program can read what you said, like “hand me that axe”, and be able to reiterate the sentence. There’s a voice lag in DDO, so during that lag if the computer could process the voice chat into that dialect.

You know what else would be cool? If your voice could be transferred into a character voice, so you’d sound like a druid or an orc, an orc accent or a night elf accent. I think that would make the game so much better – it would be amazing.

I wonder, you know how you can get those little things that change the sound of your voice? Darth Vader or … I wonder if you could put that into the game? So you could turn a male voice into a female voice?

I wouldn’t mind if DDO had some sort of filters in it to allow to perhaps sound more like a barbarian, or maybe give more of a female timbre, just to give it more character to playing the game.

Some participants enthusiastic about voice-changing software were concerned not so much with sounding like their character, but simply with disguising their natural voice.

There are people who just don’t want to be a ten year old boy. Or they don’t want people to stereotype them based on their social standing.

I hate the sound of my own voice anyway – maybe I would change it.

On the other hand, one participant pointed out that to alter a user’s voice would not merely be a matter of changing their tone or accent, but also their vocabulary, which would probably be impossible with current technology.

### 4.3.8 Sounds from physical context leak into virtual world

Participants’ concern with ‘who could hear what’ went beyond the transmission of their voice, to include sounds that originated within their physical environment. Input via a microphone meant that sound, unlike text, could breach the boundary between game and outside world.
Participants described incidents in which sounds such as breathing, eating, household noise, speech from family members, TV and music were accidentally transmitted into the MMOG voice channel.

My husband is pretty loud, and I usually play when he’s not there. But then he comes home and I’m still playing he says “What are you doing? What’s that?” I’m chatting and people say “what?”

You can turn the microphone sensitivity up or down. I had mine really sensitive because I didn’t want to talk loudly and wake up the baby. After a while I realized he couldn’t hear me, even if the baby was near me, but everyone else could hear me typing and breathing, which was really annoying for them.

There was one I think where [name] was playing classical music in the background, and every so often we’d suddenly get the classical music, the peaks of the music, coming across … And your kids in the background. Every now and then I’d hear these blood-curdling screams. [laughter]

I can’t use voice while my wife’s watching TV.

I’ve got a fairly decent headset, it’s quite good, it doesn’t pick up a lot of background noise. If I’m drinking and stuff it won’t pick it up.

Crunch crunch crunch – are you enjoying your biscuits?

DDO had a feature whereby a player’s name lit up on the screen when they spoke. This was designed to ease the difficulty of recognizing voices, or knowing who was speaking (about which more later). However it had the unfortunate side-effect of placing extra attention on users who were transmitting unintended noise into the game.

In DDO when someone’s talking there’s a green line around their character. So if someone’s breathing [into the mic] you can just look and go “Green light – who’s breathing?”

I have the mic on all the time, but I keep it a certain distance so people can’t hear me breathe. I don’t eat. [If you transmit] you can see when your [name lights up].

A participant described a situation in which he paused playing in order to go and comfort his small child, but placed his headset beside a baby-monitor speaker and thereby broadcast his efforts into the game.

I generally play with my wife out, so I get to look after the baby. Often I have to break off in the middle and say “sorry guys, I just have to resettle the baby, he’s crying” or whatever. Sometimes I put the headset down next to the baby monitor, which I also have sitting next to the computer. And I go off and say “there there” and you know, the crying and everything, and I get back and I look at the text discussion and they’ll be saying “do you think he’s winning yet? No I think he’s still battling the monster” [laughter]

Participants reported that when playing with people they knew well, they could recognize which team-mate’s spouse or child was making noise in the background.
I know [name’s] wife and I know [name’s] kids so I recognize their sound. But possibly some of the others guys [don’t know each other that well] … like I don’t think [name] has met [name’s] kids.

One participant reported overhearing younger users being told by their mothers to get off the computer. (Recordings of such events became something of a genre on YouTube.) “

We’ve had a few times with the young guys and their mother will shout “get off”, and they’ll say “Oh, I’ve gotta go, sorry guys”.

Microphones have ‘mute’ buttons which are designed to solve this problem, and MMOG voice systems could be set into a ‘push to talk’ mode to prevent unintended transmission.

When I first started to play I played hands-free. I found that when people play hands-free their mic’s on all the time, so it’s hard to tell who’s speaking, and also you get all the heavy breathing and the coughing and the snorting, screaming kids, you know, whatever. So I found it better to turn off the hands-free and use the button.

But having to use one’s hands to press a button before each utterance diminished one of the key benefits of voice: its hands-free operation. Most participants considered push-to-talk a hassle. Some were prepared to risk transmitting household noises in order to avoid using a mute button, especially if they knew their collaborators well.

Our group uses “always on”. The sociability of our group is such that we don’t mind if there’s a bit of background noise.

I don’t think any of the guys are using push-to-talk. So with some of them you get background noises coming through. … I can see in pickup groups especially that other people might find that annoying.

This clearly depended on collaborators’ offline relationships: participants who were playing with strangers were more anxious about what they transmitted.

4.3.9 Sounds from virtual world leak into physical context

Participants also noted that speech and other sounds intended to be heard within the game could be overheard by people who were co-located in a user’s physical context. Again this problem could not arise with text (unless someone were to read the user’s screen). “I think it’s important that text is an option too, for when somebody else is home and you just can’t talk.”

One participant reported adjusting his volume to achieve a compromise whereby team-mates could hear him but he wasn’t disturbing his family.
Two others reported that they played differently depending on whether family members were at home and able to hear them. In particular, if small children were around they used voice differently.

Because I’ve got kids at home, and I play late at night, I have to whisper a fair bit, and people find it hard to pick up.

Usually it’s me sitting in the corner chatting, and my wife over there reading a book or watching television a few metres away.

Just as groups who played together over time became familiar with each other’s household sounds, so the members of a user’s household could became familiar with game events.

One time [my husband] said that he could tell that they were a bunch of guys that I was playing with. He could tell by my voice that I was getting really frustrated with them, and that they were being really stupid and annoying.

The importance of this phenomenon depended on where users were located, who was located with them, the time of day, and what was being broadcast.

He [my husband] doesn’t watch TV when I’m in there [gaming].

If no-one had been in the house I might have behaved differently.

Certainly it cannot be assumed that every MMOG user is located in a physical and social context where sound leakage is not an issue. On the other hand, one teenaged participant said that he played in his room, and was not concerned about his parents overhearing him. He treated speaking on the voice channel as though he were talking on the telephone.

I just talk normally, in my own room. It’s just like talking on the phone when you’ve got that three-way chat.

Again users’ degree of comfort with this phenomenon is clearly dependent on exactly what is being said by their online collaborators, as well as their physical context and who is present in it.

4.3.10 Channel congestion

Many participants complained that voice channels were prone to congestion, far more than were text channels. The larger the group, the less useful was voice, because people became more likely to talk over each other. This confirms the findings of previous studies of turn-taking in CVEs, such as Bowers et al. (1996) and Nilsson et al. (2002).
A congested text channel, by contrast, was still usable because of the way text utterances were parcelled into discrete sentences and persisted on-screen (for prior findings see for example Brown and Bell, 2004).

It’s a very obvious observation that one of the benefits of text over voice is that people talk over each other, like in real life. In text you can’t do that. In text chat you can see everything that’s going on even if you don’t have time to keep with it simultaneously; you can refer back to what was said. There’s a history.

Voice definitely works better for small groups. With large groups you just wind up with 20 people talking over each other.

If five people type something at the same time, you can go back over it, but if five people say something, it’s just going to be noise.

WoW guilds tended to use voice mainly during raids: events in which only a subset of the guild was engaged. One participant felt that a whole-guild voice channel would be “like trying to speak in a crowded room”.

Another WoW participant reported that his group didn’t try to use voice in groups larger than 6, and felt that waiting for his turn to speak detracted from immersion.

Another reported that raid leaders, who controlled the team’s Ventrilo channel, were prepared to mute team members who cluttered it.

That does happen on occasion, when we’re in large-scale raids, like there’s a 40 man instance, and someone won’t keep their mouth shut, it’s getting in the way of the business at hand, they can get muted. You need that.

However one participant felt that text channels also suffered congestion when used by large raiding parties, and that voice was still easier overall when coordinating large groups.

Another thought that the problem of congestion would disappear over time as people became more familiar with the proper use of voice in MMOGs. “You have the usual issues when everyone tries to speak at once, but I think that resolves the more you play together.”

A participant noted that, unlike in the physical world, where conversationalists could read each other’s body language, avatars provided no visual cue that their owner wanted to speak (cf. Moore et al., 2007). “Fifty people who can’t see each other opening their mouths aren’t going to know when someone else is going to speak. It’s not going to work in my opinion.”

Furthermore the lack of directionality in the voice systems tested prevented the use of selective focus on a speaker.
In real life if you’re at a party and there’s multiple conversations going on, your ears can focus on one or the other. But [in a game] if you get more than one person speaking you can’t differentiate between them.

Participants debated the largest number of users that a MMOG voice channel could be expected to support.

I think ideally no more than 8, maybe 8 to 10 people, I wouldn’t have any more than that. I wouldn’t mind having a channel where the whole guild could chat, but the problem with that is, just as it is in a crowded room it would be very difficult to coordinate and hear people. But I can see the advantage of having a voice system that could provide something at a guild level as well.

At a small group level, voice chatting is great. Probably 5, 10 … But with our 40 man raids, there’s no way … Can you imagine 40 people all trying to talk at the same time?

A busy channel also made it harder to recognize who was speaking.

Two-way radio in ‘physical world’ scenarios (e.g. emergency response, military operations, taxis and trucks) can also suffer from channel congestion, and to solve this problem these users have devised radio discipline and a stylized vocabulary and grammar. The participants discussed whether this was or should be utilized in MMOGs. Some felt that the use of formal protocol was rare. One said that he didn’t use and didn’t want strict military-style voice procedure: this would be too formal and not fun: “what if someone wants to laugh?” Another however desired and had experienced the disciplined use of voice by raiding parties.

You’ve got two main people: the raid leader and the battlefield commander. They’re the ones talking. You’ll get other people talking – someone will say “loot drops”, and someone will say “Oh, I’m interested in that”, and then you type in your bid for whatever item, to confirm that you’re interested. But it’s never the bullshit. There’s no chit-chat.

Highly cohesive groups had the discipline to prevent voice channel congestion.

It depends on the group. Occasionally you’ll come across the very rare group that actually waits for other people to speak.

It’s usually fine. People take turns.

But network lag made congestion and faulty turn-taking more intractable for trans-national groups.

Within our group, because we know each well, we work around or coordinate what we want to say, and give each other a chance to speak. I had on one occasion teamed up with people I don’t know, and speaking over each other can be quite an issue, especially when people are at a distance. So I find, speaking with American players, there’s a bit of a lag, and we tend to talk over each other a fair bit.

Asked whether voice could ever support very large numbers of users in a virtual world, one participant suggested a system that parceled sound into neat utterances, as text channels do.
You’d have to have some sort of control structure where the voice chat was saved in sound bites and these were transmitted in an organized linear path. But you can see the nightmare of something like that happening.

Participants were asked whether a ‘public’ voice channel was feasible for areas like taverns in which large numbers of players congregated to form groups, trade and so on. While many desired such a feature, others felt that channel congestion would prevent voice from ever being used for public chat.

It’s good that voice only works within the team. I like that you can’t use voice in the pubs. I like small groups of people.

With the current technology you’d get the whole problem of people talking over each other, and I don’t think you’d be able to make sense of the conversation that’s going on. If it was to the level where you could distinguish conversation it would make it a very interesting feel. It would add a level to it. But with the current technology pretty much only one person can talk at a time.

One participant reported that to address the problem of critical messages going unnoticed, WoW add-ons were available which allowed raid leaders to flash a large text message in front of all members.

The other thing that this guild uses is another mod, which broadcasts critical messages in the middle of your screen, so it’s not only not only in the text chat, you’ll get a big message saying “heal main tank”, it’ll go ‘bang’ in the middle of your screen. So they use another form of text communication as well. … So it’s all designed to make you succeed at what you’re doing. Voice doesn’t always achieve that end necessarily because of too many people talking, too much crapping on. … Voice is useful if it’s used well, with a lot of discipline, but if it’s undisciplined, people are bored and they start crapping on, and you’re dealing with children, it can become counterproductive.

Here we see (highly stylized) text, with its ability to be stored, being used as a backup for ephemeral voice broadcasts.

### 4.3.11 Is it feasible to operate multiple voice channels?

MMOGs had typically offered several text channels, each of which followed different propagation rules. For example, ‘guild chat’ is broadcast to the entire guild, ‘raid chat’ is broadcast only to the group currently engaged in combat, ‘vicinity chat’ is broadcast to users who are nearby in the VW, and one-to-one chat or the ‘Tell’ is for private conversations between two individuals (see table 1 in chapter 1). Corresponding offline voice media with these propagation rules are radio (guild and raid chat), sound in air (vicinity) and telephone (one-to-one). Many MMOG users kept several text channels open at once and conducted simultaneous conversations in separate windows. WoW also offers some text channels that
broadcast over the entire virtual world, for specialized purposes. A participant described his habit of keeping the world-wide ‘Trade’ and ‘Looking-for-group’ channels open in addition to his guild channel.

DDO did not offer the ability to run multiple voice channels: users had one, raid-group-wide, channel. I asked DDO players whether they wanted multiple channels. Some did, and suggested that the interface be presented as a simple list of screen buttons, one for each channel.

I’d have a button for one on one, a button for group, one for guild, so you just hold it down so you’re talking to whichever. So if you want to talk one on one, talk to the team leader and say “this guy’s being rude, let’s keep an eye on him or give him a warning or whatever”. Or if it’s a guild matter, you click on a guild button and talk.

When we’re in DDO we’re just a small party anyway but I can see that it would be useful to have a guild chat channel. So people can do guild chat if they’re not in a party.

But others felt it would be impossible to monitor several conversations at once.

Is that an extra thing to worry about though? If you say something and someone doesn’t reply, now you’ve got to guess what channel they’re on.

It could get very complicated I think.

Several DDO participants said they wanted a guild voice channel, and some wanted all-to-all vicinity voice in public spaces such as taverns. “I would like it if I could go to a tavern and hear somebody say, I’m looking for fighters to go into a group.” However as noted above, they recognized that voice in public spaces could become a clamour. One suggested that the solution to this was directional sound.

Voice chat needs to get sophisticated enough so that you’re walking through a market place or something, and you’re like this situation here, where we’re sitting in a room and you can have several conversations going on at once, and you can focus on individual conversations without the other conversation just wiping it out - because of the direction of the voice.

(This was also noted in section 4.3.10.)

Some WoW guilds set up their own voice systems, and used multiple channels. A WoW user described his guild’s simultaneous use of several Ventrilo channels for raiding, officers and chat, as well as a channel for each character class.

We have a raid channel, and an officer channel for the officers to talk. So if we go on a raid we’ll get moved into the raid channel, and there’s general chat, and other stuff.

It’s quite common for the officers to go to the officer’s channel, while they discuss issues, and then they’ll drop back into whatever the raid channel is.
Another participant’s guild conducted highly coordinated raids with one “general” in overall command, several officers, and a leader for each class.

In WoW I raid in a group of 40, so you’ve got different channels, you’ve got the raid channel that everyone’s on, and the class channels, like warriors will be on one and druids on another. It’s not a good idea to monitor both at the same time, because you’ve got a class leader, the warrior who’s leading your class saying “your character needs to do this”, so you just do it. Because you’re doing that, the other people don’t need to know what you’re doing because their class leader’s telling them what they need to do. So the priests’ class leader will be saying “we’ve got [name] tanking one of the guards, we need priest A and priest B to keep him healed, if [name] goes down we need druid A to be able to battle”. If you’ve got everyone able to access the same channels you’ve got battle confusion.

Golub (2010) described a similar organization in his WoW guild.

Some WoW participants reported embarrassing situations caused by a user typing a message into the wrong text channel.

I know in WoW you do occasionally make mistakes. I know I’ve been whispering to several people and … you say something in guild chat that was party chat or whatever.

I’ve had instances where I’ve been in a party doing a quest and role-playing with them [in text], and I’ll also be separately role-playing in a private chat with another friend. I’ll be sending Tells or Whispers to an individual person, where I’ve had three separate conversations going on, and I’ve sent the wrong thing and they’ve responded and gone “what?”. Which has been quite embarrassing especially if you’re bad-mouthing someone or just talking about something completely different and if they take it out of context, what you’re saying. I can see the benefit of it, being able to talk to someone individually [in voice], like being able to send messages to individual people, but …

This participant hints at the problems with presentation-management that might occur if users had to manage simultaneous conversations in several voice channels.

Another participant felt that transmitting into the wrong channel would be worse in voice: “Sometimes you’d hear things you didn’t want to hear, or people who were annoying because they were having a long conversation right next to you.”

Some participants wanted to be able to speak one-to-one with other users, imitating the ‘whisper’ or ‘tell’ text channel. Example scenarios were negotiating while trading, and dispersing loot after a raid, during which issues could arise that needed to be discussed without the whole group hearing.

Voice doesn’t allow you to be as personal as to talk one on one. You can only still talk within your party.

I think the whisper is important too. There are some things you want to talk about that you don’t want to subject everyone to.
There are times when I privately want to talk to someone. As an example, I knew there was an item that I’d pick up, which I wanted to pass on to one particular person, I didn’t want to broadcast this to the rest of the group. We have an understanding that we usually open it to the group, whether they want a particular item or not. I didn’t want any resentment or hostility. I knew the person was actually after it, and had had back luck after bad luck getting the item, and I simply wanted to offer it to them. I had to do it via text. It could be done better, a lot quicker, by using voice, because we were actually hurrying through the level, and it was hard to actually get to a point where I could do that.

You should be able to target an individual person. You already target people in a Tell by clicking on them. If I’ve got them in my cross-hairs, why can’t I click something else that says “talk to them”?

You could move your mouse cursor over an avatar, hold the middle button and speak to them. Then if they want to respond they could also hold their mouse button over the toon and say “you’re welcome, any time”.

Some participants who were enthusiastic about voice wanted to be able to speak with anyone they encountered in the virtual world – not just pre-defined team-mates.

I would like voice in WoW where you can talk to anyone in the game, not just guild, so if you go up to them and say “Hi how are you going” you’re talking to them – rule out text chat altogether. I’d love that – it would be great. You’d be able to talk to the Horde, taunt the Horde, shout at them.

Participants discussed the pros and cons of visually representing instances of one-to-one talk as avatars whispering to each other.

In some ways whispers could be appropriate, because I could walk over and whisper in [name’s] ear, and say “so and so’s a bastard”.

I wouldn’t want my avatar to have to do that. To have to walk up to someone and lean in and cup my hand around their ear.

A participant suggested that ‘tells’ be implemented such that senders could address recipients simply by speaking their name, similar to a feature offered by some mobile phones.

It would be interesting if you could do those direct Tells not by targeting them but by calling them by name. You’d have to have some sort of voice recognition.

To avoid the possibility of abuse, they suggested that such mechanisms represent a “request to chat” rather than automatically opening a channel.

4.3.12 Use of both text and voice

Some participant reported playing with groups in which some of the members used text and the rest used voice. These groups tended to split into two sub-groups depending on channel, so that it became impossible to coordinate the whole group.
Generally the poor text people get ignored when we’re in the middle of a battle or something. A couple of times it’s happened that a poor text person has gone “help” because they’re dying, and I’ve noticed that and said “hey, does someone want to heal [name]”.

The rest of the group was using voice and I wasn’t, so I was pretty much at a loss for what was going on.

However in some cases conversations flowed between the two channels.

Sometimes you’ll even have it spanning both – the same conversation going on in both text and voice.

I’ve had conversations go on in both, like switching between voice and typing, back and forth, the same people, they’ll say some things and type some things.

Others, after starting to use voice, stopped checking the text channel.

Text usually is something we use when the voice chat breaks down, or we’re communicating with people outside our group. Maybe for emotes. But I’ve got to the point now where I won’t look at it unless I know someone hasn’t got their headset working and we need to watch them or we’re going into a public area or something like that.

Some voice-using groups continued to use text as a backup for important messages, or for situations where voice wasn’t working for some team members.

I find in my guild, in which it’s compulsory to use Ventrilo when you’re doing these raids, they still use text to convey essential messages, so that there’s no possibility that the message is not being conveyed. Because the raid leaders and officers – if you make a really stupid mistake, and they get really crapped off, and for good reason – you might spend three hours on a Saturday afternoon trying to get to an endpoint, and if someone f***s it up because they’ve done something stupid. So they make it quite clear in text that these are the dos and the do-nots, as well as chatting about what they’re doing in the voice client.

There’s no reason to use text at all if everyone’s got voice. Except in some situations, like if I have to leave, and everyone’s busy, they’ll miss it, so when I’m about to leave I just type “AFK” and that’s a record, so when they go “hang on what’s going on – oh ok”.

Having a mixed text-and-voice team required that some messages be delivered separately in each channel, creating extra work.

If there was someone who wasn’t using Ventrilo they couldn’t take the commands the same way, so you still had to communicate with them by text – so it actually took longer anyway. We’re not on the same wavelength.

In this situation some teams defaulted to text.

Not everyone’s in the voice chat, so if you want to say something you want to everyone to hear, you say it in the text chat.
Some groups used both voice and text because some members were willing to receive, but not transmit, voice: “We got them to enable voice so they could at least hear us talk, but they would have to type to ask us.”

Some simply used text as a way to set up voice communication.

The times I’ve used text is when I’ve entered the game and to let people know that I’m in. DDO doesn’t have a particularly good way of letting people know you’ve come into the game. By typing in that I’m there, as that character, they can then invite me, and then we use the voice.

Despite some problems using multiple modalities, the use of text to debug voice and to act as a backup when voice failed suggests that a voice-only MMOG is unlikely.

4.3.13 Asynchronous communication

In a persistent virtual world it is likely that only a subset of a guild may be logged in at a given time, while activities such as planning and coordination need to be on-going. This implies a need for asynchronous communication. WoW players reported that some guilds were solving this via applications outside of the game software. Text forums and email were the preferred media for asynchronous communication between game sessions.

However the asynchronous nature of text communication could be a benefit in real-time communication as well. Use of multiple text channels is feasible partly because users can scroll back through message logs to re-read parts of a conversation that they missed while attending to another channel. (These are essentially asynchronous even if they are used synchronously.) Participants reported a frequent need to scroll back, such as when they came in mid-way through a conversation or had taken a break from the game.

Most participants felt it would be impossible to implement a ‘scrolling’ mechanism in voice. Voice could perhaps be recorded, but it would be hard to scan through it.

Theoretically you could have it record the voice and scroll back, but I think it would be too hard. You can’t scan across voice like you can text. Unless you had a record of each contact that was made, that said something like “Jim said something at this time” and it played back what Jim actually said.

Some felt that the ability to scroll back was a significant advantage for text.

In text chat you can see everything that’s going on even if you don’t have time to keep with it simultaneously; you can refer back to what was said. There’s a history.
When everyone’s speaking at once, with voice it’s gone, whereas with text you can scroll back and see it.

Sometimes you might put your headset down to go to the loo or grab a cuppa, and when you come back after a few minutes you miss out on what’s been going on. Whereas with chat you can scroll back and read it and see what’s going on.

Participants were asked how a voice channel might be designed to work asynchronously. They discussed the possibility of emulating answering machines and voicemail.

You need voicemail! You click a button to say you’re not available, you’re on a quest or something … “I’m having my head removed by a dragon. Please leave a message” [laughter]

Answering machine! “You’ve got three whispers waiting for you”

I can foresee having in icon pop up saying “someone has sent you a voice chat, do you wish to play it”. Like an answering machine. So when you’re done with the battle, you press the icon and you can decide whether it’s important or whether it’s BS.

You could set up your macros automatically so the moment you go into combat, automatically all tells go to voicemail.

Some suggested that a ‘virtual answering machine’ thus implemented could be used as a way to deflect messages while busy, much as some people use voicemail to avoid taking a phone call.

You know how you’re getting a tell – you hear the whisper. Whether you accept it is another point. But there’s a button and you can replay it. Or you get a whisper and you can send it to voicemail. You know how you get your mobile phone and you go ‘reject’.

Some participants suggested being able to “time shift” the voice channel in the style of TiVo.

Even if you’ve got a buffer. Part of your client that sits on your PC.

Like time-shifting for TV?

Yeah. You “pause the show” as such. Then you can come back and replay it.

However participants were apprehensive about the thought of having to check for and play back messages. One thought it would be simpler just to ask the sender to send again.

I don’t think we’d do that though. Think about it. Say you’ve just logged on, and I was going to send you a voice whisper. You wouldn’t want that to be just coming through, because if you hadn’t plugged in your headset, your kids are going to be hearing “Hey [name] how are you” coming through your speakers and you’ve gone to the toilet or something while it was loading.

If you knew it had come from Jim, you’d just say ‘Jim, what did you say back there?

Another suggestion was that users who missed a voice utterance be informed of this by a text message on screen.
Why can’t it generate text for you? “You have received a voice tell from blah”. So you connect with
them and say ‘sorry I missed it, I was in a battle’.

One WoW participant described problems receiving voice chat from guild-mates while in the
heat of battle:

All of these situations where you can’t respond immediately, or it’s too confusing to have people who
aren’t there in your party talk to you while you’re in a quest, that’s where text comes into its own. You
don’t mind using text for that kind of thing.”

It was suggested that in this situation, the game should signal a user’s unavailability by
responding to voice calls with a canned recording of battle sounds.

4.3.14 Integrating voice with the virtual world

The third-party voice products used by WoW and Everquest users were not integrated with
those virtual environments. Their users’ ability to speak with each other was not affected by
their positions or actions in the VW. They could even log out of the game and continue
speaking with people who were still logged in. Channel control – specifying who could speak
with whom – was entirely up to the users themselves to administer.

By contrast, DDO’s voice system was integrated with the game, enabling a number of specific
features. DDO automatically connected the voices of users who had formed groups. It
displayed visual cues on-screen about use of the channel, highlighting avatar names when
users spoke. This eased the problem of matching voices to avatars which affected non-
integrated channels.

The thing I think looked great about it was … on the screen you could see the icons for your group,
and it highlighted whenever someone was talking. I think that’s a really great design feature …
because that’s one of the issues I have with other voice clients … you can’t necessarily tell who’s
talking.

However this cue was not completely successful. One participant noted that she couldn’t
always see the small indicator, and that looking for it diverted her attention from other parts of
the display.

Maybe it’s because the writing is so small or my eyes are so bad, but I don’t necessarily equate the
voice with the character name. I just take it that it’s one of the team members speaking and I don’t
know which one. It matters when someone says “who got the plus one sword” and I have to ask “who
said that?”.

Also, the visual cue disappeared immediately the person stopped speaking, so that if a user
happened to glance away during an utterance they would find themselves wondering who had
spoken it. To solve these problems, some participants suggested that whole avatars should light up, or that a speech bubble should appear (cf. Brown and Bell, 2004), or that the sound of a voice be directional (like the system examined in the next chapter).

When people are talking they should have a little bubble coming out of their mouth. You see a thing on the side [of the screen, name lights up] but you’re looking at something else.

I want the little on-screen avatar to demonstrate that it’s the one that’s speaking. Some kind of voice bubble or it glows or something. They [the character names on side of screen] flash, but I’m not looking there, I’m looking here [at avatars in action]. I’ve got a wide screen and I have to deliberately look at those things. I usually want to know who it was after they’ve spoken as well, and then it’s too late because they’re not speaking any more. And also when people are speaking simultaneously. You could have a record! You get a record with text. You could add a line in the text window with the name of the people who speak. Obviously you couldn’t say what they said, but you could just have their name and “said something”. I don’t mind if I lose what people said, but I want to know who said it.

Non-speech noise also caused speaker names to highlight in DDO, causing embarrassment as discussed earlier, and reducing the power of the on-screen cue. “I found that when people play hands-free, their mic’s on all the time, so it’s hard to tell who’s speaking.”

One participant noted that in a pick-up group, one is meeting several people for the first time, all at the same time. None of these people would know each other’s voices, and they needed some kind of help from the software to negotiate this social complexity. They felt that highlighting avatar names was not enough.

You’ve got the character screen and you can see the light go on when they’re talking. Sometimes I can recognize the voice, sometimes it’s difficult when two people sound the same. I’ll give you an example. If you’re grouping with someone from the east coast of America, and someone from Australia or England, and they have a very distinct dialect, and then you get three people from down south of America, and you’ve got a couple of blokes sounding like good old boys, they’ve got similar tone and inflection, it’s so hard to tell the two apart.

A DDO participant noted that users didn’t address each other using their real names but used in-world character names, causing some confusion.

Then it’s “Hi, Elle” – it took me a while to realize that was me. Everybody says “hi”, but you don’t know who everybody is.

Participants sometimes struggled with how to pronounce these character names, which consist of syllables in a fictitious, Tolkien-esque language. One participant heard users shortening or familiarizing unwieldy names. Another was uncertain at first how he should pronounce game-jargon and acronyms.

Some things you type all the time, but you never actually say out loud. The first time I go to say it, I stumble over it. Do I say ‘exp’ or ‘X. P.’ or … .
One participant felt that indexical speech didn’t work when a user was playing with strangers whose voices they didn’t know, because it was insufficiently clear from which direction a voice was emanating. These issues are addressed in studies two and three.

4.3.15 Not knowing who was speaking

Most participants reported problems identifying the speaker of an utterance. This is not a problem with text communication, because utterances are prefixed with the sender’s username.

You need some way to know who’s talking.

At the moment there’s a real disjoint between what you hear, what the avatar’s doing and looks like, and what the text thing is. So at least with text you’ve got the name of the character to link it [to what they character is saying]. With voice you don’t get any connection to the name of the character or to the character itself.

As discussed in section 4.3.14, DDO highlighted players’ names on screens when the spoke into the integrated voice channel.

However third-party voice clients such as TeamSpeak, used by WoW and Everquest players, were not integrated with the games and couldn’t highlight the speaker’s name.

That’s one of the issues I have with other voice clients … is that you can’t necessarily tell who’s talking.

In raids, the fact that TS is a separate program means that you don’t have any way in the game of knowing who’s speaking, but you get used to knowing whose voice is whose. Unless you know their voices there’s no way of knowing who the character is who spoke. Especially if you’ve got 24 people in a raid, somebody talks … You can work it out sometimes by what people call each other.

Participants suggested extra features which might assist the user in identifying speakers. “Something I’ve really wanted after using it is something appearing over the character’s head while they’re speaking so you know.”

However not knowing who was talking was not always a problem, and announcing oneself before each utterance (in the style of military radio operators) did not appeal to participants.

Sometimes you don’t care who it is. You’re chatting to someone, and someone in the group says something, in the heat of the moment it might not really matter. What if someone wants to laugh, do they have to say “this is Charlie?”

This disinterest in strict voice protocol was universal among participants – something I found surprising since the games they were playing were to an extent simulations of infantry combat.
4.3.16 Voice and the ‘atmosphere’ of a virtual world

Although most participants appreciated the advantages of voice, some felt that an MMOG lost some of its atmosphere when users could hear each other speak.

You lose the magic to some extent.

Voice didn’t feel right. I was used to typing.

It was a little off-putting. It’s a different communication style. Suddenly going from reading and typing to actually having to listen to people talk.

Perhaps I’m not as outgoing as some people but I don’t like to hear them. With the text you can play with any age group. But if you’re a 32-year-old business analyst, maybe you don’t want to hear 14-year-olds giggling on the other end.

Just as one participant described herself (above) as disliking text-based communication, some participants said there were MMOG players who simply don’t like voice.

There’s usually only about 10 people on Ventirlo because not everyone uses it – some people don’t like it. Some people would rather not talk to people.

These data indicate that for some users, part of the appeal of virtual worlds is the limited social presence of other people. Users want there to be other people in the world, but don’t want them to be too clearly perceptible – as people at least. This may be an example of the ‘alone together’ principle of MMOG use proposed by Ducheneaut et al. (2006), However this probably doesn’t seem to apply to MMOG players who group with friends.

4.3.17 Sound quality and glitches

“Technical glitches” are mundane problems. Yet all participants experienced problems with sound quality at some point, so much so in fact that many felt that sound quality problem significantly influenced the user experience of virtual worlds. Almost all of the DDO players had trouble getting voice to work at first, though most were satisfied eventually. Participants complained about complexity and a lack of assistance for new users.

The built-in voice aspect looked great to me, but I couldn’t get it to work. I’m not sure why. People could hear me, but I couldn’t hear people. I think it was a settings issue.

I had gain the wrong way round. So I kept turning it down, and people could hear me less. That took a few weeks to get straight.
I couldn’t even work out what you had to do. There was nothing in the manual that said you press this button to talk, to activate your microphone, or you just join a group and then you could talk, or what. So I never used the voice, because I couldn’t work out how.

Setting up DDO voice seemed to be easy, but inexplicable things would happen. If you disabled and enabled it again it would suddenly work. There was a lot of fiddling. And we’re not manual readers. There are slide scales and volumes, things like that which we weren’t really sure of. We’d fiddle with the slide bars and the person would say if it was clearer.

Some of the DDO voice problems were evidently server-wide, and eventually fixed by the vendor.

Early days there was a lot of complaints in the forums. There was a lot of stuff about working on issues, which they eventually seemed to resolve.

At the beginning there was a week where there was so much lag you couldn’t use [voice]. That was quite annoying.

However other problem seemed to be a permanent fixture of voice use.

Glitches in voice transmission could become significant usability problems which made the game almost unplayable. If one group member had a communication problem, all members were affected.

My voice always works, but I reckon most people don’t know how to set it up properly. Most people have got voice, but it breaks up. A couple of times I’ve had people who were saying something and they’re like [makes sound of vocal breaking up] and you’re like “oh I’m sorry I didn’t catch that, can you repeat it”, and they’re like [noise again] … “what?!”. 

Users of third-party services had more success than DDO players. but reported some problems.

In EQ it was very straightforward. I saw someone had set up a server, downloaded the latest client and installed it, I didn’t need to use help or anything, I just set it up, connected to a server, and started talking. Well almost. I had a few problems with the microphone. I eventually got it working and then it was fine. I was amazed how well it worked. It was plugging the mic into the right slots. I was able to set it up so the EQ launcher automatically launched it. I play EQ in full screen mode, and I just hold down a button to talk.

The Vent server’s gone down a few times, but luckily not while we’re in a raid. Then we’d just have to revert to raid chat. That would be annoying.

Problems with voice were not always a matter of the system being down. Delayed transmission (lag) was a frequent problem. It seemed people are more sensitive to lag in voice than in text, and that it causes a breakdown in conversational protocol.

There were a number of bugs. There was quite a bit of lag. There were compatibility issues with some of the equipment, and I had to go out and buy a new set of headphones because it just didn’t work with the ones I had. Once those issues were resolved it was fine.
With DDO you’ve got close to a sentence lag time. So you’ll start to say something, and then stop because someone else is talking, because they didn’t realize, just because of that time delay. Then you’ll stop because you’re being polite. Then at that stage they’ll hear your voice, so they’ll stop. So you get these jerky sentences back and forth until someone just says “oh screw it, I’m just talking over the top of everyone”.

DDO players reported that at the start of each play session, the members of a group needed a coordination phase during which the status of each user’s voice setup was adjusted using spoken or typed feedback from team-mates. This process was more arduous in pickup groups, whose members would usually be unfamiliar with each other and the quirks of each other’s hardware. Participants reported sessions in which they assumed all along that they either could or couldn’t be heard, only to find out later that their assumption had been wrong.

Many of the DDO users that participants encountered early in the study appeared either not to have voice hardware, or to be unable to use it, or to have it switched off or incorrectly configured. Some participants reported spending their first couple of play sessions frequently saying “Hello, hello?”, unsure whether other users could hear them. (It seems that pretending that one’s microphone isn’t working has become a recognized strategy in VWs for users who do not wish to use voice.)

The ‘chattiness’ of groups encountered by our participants varied. Some participants joined pickup groups in which some users would rarely speak, though they were apparently listening to the voice channel and might comment on voice utterances using text.

It’s more like certain people have a bad connection or bad equipment or something. Like there was one guy the other night, and everything he said, it was like the game was turned up too high or something. I couldn’t understand what he was saying.

_They can hear you fine._

Yeah, and some people are crystal clear. And then the other night someone was having problems with their microphone jack. There was static coming over every now and again. 15 minutes later it was still happening and I was like “dude, turn off your microphone”.

Lag led to moments of comedy as conversations became hopelessly disorganized.

On Saturday night we had a bad problem. It was very funny. I fell out of sync with everybody else by about two minutes voice-wise. I could hear them in real time but anything I was saying was about two minutes delayed. So I was talking to [name] on a mobile and then two minutes later he was laughing because he could hear me responding to him on the mobile, coming over the voice.

You’ve got that ladder that goes down [in one of the DDO dungeons]. So we’re trying to direct him [a team-mate] via voice. He’s almost on the ladder and then he goes off to the side. We go “no no no stop stop!” and he kept on walking, and then I’m “ok forward forward forward”, and then we’ve got him lined up, and then he started to back up, and for some reason he went off-centre a bit, and so he’s
backing up and backing up and we’re screaming “stop! stop!” It must have been because there was 

enough lag there, he didn’t hear it, so he just went straight to the bottom of the pit.

Getting loudness right was the most common problem with voice. Loudness has no analogy in 
text and makes voice an inherently complex medium. Some participants commented that they 
wanted to normalize the volumes of individual team-mates, but the software did not provide a 
way to do this.

Different people come through at different volumes. I can only adjust it for the group. So if I adjust it 
up, someone might be really loud whereas I can now just understand [the quiet person]. So sometimes 
you have to say to someone “adjust your volume up”. I think it would be good if you could 
individually adjust the volume up for that individual person. If you’re dealing with a group of six or 
more it can be quite fiddly due to different people having different setups. Sometimes we have to 
pause while people say “Can you hear me now? Am I breaking up?” Whereas the last time we were 
using exactly the same settings and they were fine for the group.

The most annoying thing voice-wise is when they’re too quiet. They’re talking but they’ve got their 
volume set or tolerance set too low so you can barely hear them.

It took me a while to get my microphone set up. I thought the volume was my volume. At one stage I 
was told that my volume was too loud. So I turned the mic volume down, then I could barely hear 
anyone else, until someone told me “oh you need to turn the volume up”. I must have been playing for 
three weeks until a guy in a group said “turn up the mic volume, and then turn down the microphone 
gain”. So when I did that I was able to increase their volume to my speaker, and decrease my volume 
to their speaker. That was not made clear during setup.

By contrast the text channel was regarded as reliable and a useful backup when voice was 
failing for any reason. One participant described text as a ‘safety net’. “Voice has gone down a 
few times due to bugs, so falling back on the text has been very good. “

Some of these ‘technical’ problems are implementation-dependent. Others however, such as 
volume mismatch due to microphone placement, may be very difficult to solve and might be 
considered inherent to the voice modality.

4.3.18 The possibility of spatial voice

While commenting on the potential for voice in large public spaces, several participants 
suggested spatial propagation as a solution.

I think it would work with voice. If you had some kind of proximity thing it might.

You could have a proximity-based thing as well, where the volume goes up and down depending on 
how close your characters are physically to each other. So you could target one by whispering in that 
person’s ear. Or you could just talk in a room, the volume would be within a fairly close distance, 
would drop away to the background. You could walk into a tavern and everyone would be talking 
and you’d hear that background chatter that you hear, but it would be still different to when you walk
in and it’s empty. If you walked up to someone and they were talking out loud you’d start to hear what they were saying, and then you could join in. You’d probably want lips moving or a voice bubble when someone was talking so that would help you identify.

However some participants foresaw that the benefits of spatial voice might not be as straightforward as they appeared at first glance.

It might be important to know the difference between someone standing next to you and just saying something and someone actually standing next to you and targeting you, because if you target them and say something, they might assume that their other friend who’s standing right there can hear it but they actually couldn’t.

Even in text, people, when they’re around in that environment, they might talk out in the vicinity chat, but then when it’s appropriate they might move to private group chat or Tells or something. Then there are other people who just don’t have that presence of mind and talk just in the vicinity chat even when they really shouldn’t be. I think you’d find with voice, most of the time it would be all right, but sometimes you’d probably hear things that you didn’t want to hear or, annoying because they were having a long conversation right next to you.

My concerns would be, how do you regulate it? How do you regulate the people trying to have a decent conversation against people just trying to be an idiot? I want a function to be able to turn off that idiot. You obviously can’t have it in real life but if you can have it in the gaming environment

Spatial voice was not available at the time of study 1, but is addressed in study 2.

### 4.4 Discussion

Conversation within MMOGs presents a fascinating case with potential to extend the scope of research into mediated communication. The combination of role play and team coordination in a persistent simulation of fictional space provides a unique context in which to develop theory on media preferences and the influence of medium, context and user characteristics on communication.

The aim of this study was to answer the research question:

**RQ1: How does voice influence the user experience of massively multiplayer online games?**

Rather than survey a large number of MMOG users, I chose instead to undertake intensive research into the experience of a relatively small number of people. This approach traded off richness and depth of data against the generalizability of the results.

The richness of the data is apparent in the complex scenarios these participants related when discussing how using voice influenced their experiences. The risk is that with only 15 informants, the findings might not represent the population of people who use voice in MMOGs, or who use it in MMOGs other than the ones examined here, or who might adopt
voice but have not yet, or might use an MMOG but have not yet. I mitigated this risk by assigning different participants to different conditions, attempting within the time and resources available to cover the diversity of conditions relevant to voice in MMOGs. Some participants played an MMOG with integrated voice, while others used popular third-party voice products in different MMOGs. Of the former cohort, some played with people they knew, while others did not. The participants varied in their prior experience with MMOGs and with voice communication systems for games.

Diversity is a key theme in the results of this study. That is, the diversity that exists among users and their goals and contexts drives diversity in the utility of voice. Since diversity was detected even in this small sample, it seems likely that it prevails in the population as a whole.

4.4.1 The variation in use-cases

I found that MMOGs are socially complex virtual environments that present diverse communication contexts that must be negotiated by users. Participants described activities as diverse as exploration, trade, negotiation, team management, combat and role-play.

Supporting the taxonomic work of Bartle (2003a) and Yee (2006), I found that different people are using the virtual world for different reasons. For example, some like to role-play fictitious characters, others enjoy team-work and socializing, while others focus on succeeding in the game. Some are collaborating with people they know offline, others with people known to them only within the game, still others with strangers they are meeting for the first time. These conditions impact people’s choice and use of modality.

MMOG use takes place in a variety of physical-world settings, and these too influence people’s choice and use of communication medium, confirming prior research on the importance of context in technology use. Supporting the thesis of McCarthy and Wright (2004), I found that the MMOG is used very much within the context of home and family life. And confirming the findings of Yee (2006) I found that MMOG players are not all teenagers playing in their bedrooms - though some are – but include old and young, singles and couples and parents of small children, whose context of play involves family members, house-mates, work-mates and so on.. Whereas communicating by text is mostly unaffected by the user’s physical context, the utility of voice depends heavily on where the user is and who is around. Voices cross from the game world to the physical setting and vice-versa. In fact voice channels connect the physical contexts of the users on the channel, so that MMOG voice users
effectively become part of a large media space. As was found regarding mobile phone users (Madell and Muncer, 2007), MMOG users are aware of the possibility of being overheard, and tailor their choice and use of media to suit this contingency.

The results of this study indicate that there is a paradigm communication scenario in virtual worlds to which voice is ideally suited. This occurs when small groups of people, who know each other and are comfortable speaking with each other, are engaged in fast-paced activity which needs to be coordinated synchronously. The people in this paradigm situation are not physically co-located and may be somewhat dispersed also in virtual space. They are also not located in physical settings where the overhearing of conversation by co-located others is a concern. This is the situation most MMOG players encounter while raiding (Golub, 2010). In this paradigm scenario we see the benefits of voice predicted by classic media richness theory come to the fore, as this modality affords dramatically improved coordination over text.

This is also the scenario faced by most users of networked first person shooter games (Reeves et al., 2009). Players of these games have enthusiastically adopted voice communication. They encounter situations which are analogous to those faced in the physical-world by groups of emergency workers, convoy drivers, soldiers and so on.

Offline groups engaged in analogous activities choose radio for its ability to deliver spoken utterances instantly to any point in space. Two-way radio has been the medium of choice for such groups for many decades and it would seem that it is ideally suited to the dynamic coordination of a small group moving through space. But the users of both physical and virtual radio systems face similar problems of channel congestion, eavesdropping, and misidentification of voices. Physical-world radio users have devised voice protocols that ease these problems; however these do not seem to have been widely adopted by videogame players.

Groups whose members know each other and play together primarily to raid successfully, rather than to role-play or explore a virtual world populated by strangers, enjoy the voice channel’s support for rapid communication and resolution of ambiguity. A third or more of MMOG users are in-world to maintain real-world friendship groups (Williams et al., 2006). This was true of one of the groups in this study. For such users it appears that the greater social presence of voice communication relative to text enhances their experience, and it allows them to enjoy the easy sociability enjoyed by the co-located players of LAN games (Taylor and Witkowski, 2010) and multi-player consoles (Aarsand and Aronsson, 2009).
However, people whose team-mates are not known to them offline may prefer the social distance of a text channel. And insofar as other aspects of MMOG game-play do not resemble this scenario, the utility of voice diminishes.

4.4.2 Comparing text with voice

Figure 9 summarizes differences between voice and text in the MMOG context that were discovered in this study.

My use of multiple dimensions is based on the approach of Media Synchronicity Theory (Dennis et al., 2008) which analyses communication media according to the dimensions of transmission velocity, symbol sets, parallelism, rehearsability and reprocessability. To these concepts I added new criteria that arose out of the study such as the fact that voice transmits more information about the user, is prone to eavesdropping, is more prone to technical problems, and is felt to be more suitable for groups whose members know each other.

Properties of text vs voice media

<table>
<thead>
<tr>
<th>Text</th>
<th>Voice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transmits info about users</td>
<td>More</td>
</tr>
<tr>
<td>Low fidelity, difficult setup</td>
<td>More</td>
</tr>
<tr>
<td>Prone to eavesdropping</td>
<td>More</td>
</tr>
<tr>
<td>Supports asynchronicity</td>
<td>Less</td>
</tr>
<tr>
<td>Store and search messages</td>
<td>Worse</td>
</tr>
<tr>
<td>Supports group coordination</td>
<td>Better</td>
</tr>
<tr>
<td>Better if users know each other</td>
<td>More</td>
</tr>
<tr>
<td>Parallel conversation threads</td>
<td>Hard</td>
</tr>
</tbody>
</table>

Figure 9: Text vs voice in MMOGs
Transmission velocity and symbol sets are wrapped up in the dimension of ‘supports group coordination’.

MMOG players appreciated the ability to store and search text messages (reprocessability) and discussed whether this could be supported by a voice channel. They also appreciated the ability to conduct multiple parallel conversations.

Voice’s lack of rehearsability was mentioned by players when discussing the flaming that could erupt when people spoke before they thought.

### 4.4.3 Comparing different voice configurations

My study design allowed for a comparison between the voice system integrated into DDO and the non-integrated third-party systems which at the time of the study were the only option for WoW and Everquest users. Integration of voice with the virtual world of DDO enabled two features:

- when a user spoke their name lit up on screen, and
- membership of a radio channel was automatically provided to users who had nominated themselves as a group.

Non-integrated voice provided a different experience. First, there was no indication within the game visuals of who was speaking. Since this and previous studies indicated that gamers have trouble identifying voices, one could conclude that DDO’s visual cue is valuable. However a number of DDO users said that the details of its implementation reduced its usefulness: in particular, if a user was not looking at the right place on the screen at the moment someone spoke, they would miss the cue and wonder who had spoken. To solve this users asked for the cue to slowly fade, or for speech bubbles to display above avatars.

The second difference was that third-party voice users had to do their own channel management. Although this obviously requires extra work, none of our WoW and EQ participants complained about it. The reason might lie in the way groups themselves were arranged. DDO did not specifically support long-term guilds. Users simply logged in, grouped for the current session with whomever they could find, and proceeded to conduct a quest, and DDO’s voice system supported this in a clear and direct way. However some DDO users commented that they would like to speak with others before grouping with them. The process
of forming groups itself requires negotiation and judgement of the worthiness of potential comrades, during which voice would be useful.

In WoW and Everquest, on the other hand, the most important social structure for most users is the guild (Williams et al., 2006). Guilds are long-term groupings. A raiding party will usually be chosen from one’s guild, and it is through the guild that the administration of voice channels is conducted. Most WoW guilds are apparently quite efficient at self-management, and can obtain value from the opportunity to fine-tune channel configuration. DDO’s automatic voice channel management did not provide a large advantage over the third-party products used in other MMOGs.

Third-party products were actually advantageous in one respect: the channel operated independently of who was logged into the game, and even of whether the game was working. Some MMOG players reported using third-party VoIP to arrange strategy prior to logging into the MMOG, and during the down-time caused by game server crashes.

Most of the problems discussed by participants, such as channel congestion and exposure of identity, applied to both types of voice systems, and the features that users suggested be implemented were relevant to both. Many users expressed a desire to speak with passers-by in the virtual world: i.e. people outside their defined groups. They wanted to be able to set up one-to-one voice conversations, analogous to telephone calls. They were concerned that the lack of a voice log or transcript made it more likely they would miss messages, and they were concerned about broadcasting utterances to unintended recipients in their virtual or physical environment.

### 4.4.4 Voice and identity play

In contradiction of the assertion that role-play is a widespread, fundamental activity in MMOGs, most of the participants in this study denied being role-players - in the sense of voice-acting one’s on-screen character – and felt that this behaviour was rare. However there was considerable debate among the participants as to what exactly ‘role play’ was. Many felt that the sense of ‘voice-acting’ just referred to was too narrow a definition, and that instead role-playing meant playing a role in a team, or becoming immersed in the fictional setting of the game. By this broader definition of role-play, voice would not be the obstruction that many apparently have felt it was.
Although most denied engaging in role-play, many participants had encountered and/or engaged in gender-play – using an avatar of the opposite gender - and many reported that voice could make this an uncomfortable experience for everyone concerned. In fact this was one of the most-aired criticisms of voice in this study.

But it was interesting that the degree of discomfort people experienced with gender mismatch seemed to depend, as so much else did, on whether the people concerned knew each other. Furthermore, participants who had played pen-and-paper role-playing games reported no discomfort adopting arbitrary personal characteristics in those offline situations. This suggests that the mismatch that causes so much discomfort relates to visuals, and surprise. When role-playing around a table, no-one changes appearance and everyone present knows who is playing what character. But when people encounter strangers in a virtual world, they have only avatar appearance to judge them by – until they speak.

Many participants in this study felt that the solution to this mismatch was voice-changing technology. However it should be noted that few of them had actually tried this. Prior research into the Xbox Live channel, which supported voice-alteration (Gibbs et al., 2004), indicated that in that system at least, players didn’t like altering their own voices and didn’t like it when others did it. I conclude that no clear solution to the problem of gender and voice surfaced in this study.

### 4.4.5 Could an MMOG offer only voice communication?

The release of the voice-only Xbox Live network in 2003 raised the possibility that vendors might want to release voice–only MMOGs. A question worth asking then is: can an MMOG be operated using only voice communication? One could infer from this study that the answer is: “not without difficulty”. This is compatible with the finding of Williams et al. (2007) that voice plus text (not voice alone) was the optimal communication arrangement for groups of WoW players.

Some of the preconditions for a voice-only MMOG would include that the voice channel must be more reliable than existing systems, and that user setup and operation must be foolproof. It appears that MMOG players use text channels to debug voice channels.

Second, a way would have to be found to run multiple simultaneous voice channels. While some high-level WoW guilds used multiple channels, it is not clear whether the degree of discipline and organization required to do this would be found in novice or ad-hoc groups.
Third, users would have to accept the loss of *social distance* that text provides, with possible implications for role-play, gender-play, interacting with strangers and so on. Study 1 indicates that MMOG players would have mixed feelings about this. Transmission of personal information about players and their settings was broadly accepted only within groups who knew each other.

Finally, users would have to accept the possibility of collaborators in the virtual world and friends and family co-located in their physical context overhearing each other during their (typically long) MMOG sessions. Again, only among close groups of friends was this considered desirable by participants in this study.

### 4.4.6 Recent developments

Blizzard has recently implemented an integrated voice channel in WoW. Anecdotal evidence suggests that it is inferior to third-party products and has been rejected by most users. In a recent interview, a Blizzard developer admitted:

> the Voice Chat we included is not a good feature. We tried, we made some mistakes, and we accept that it isn't something used by a lot of players. It is on the wishlist of things we would eventually like to improve. A big thing it is missing now that we would like to fix is the ability to communicate with people before you get into the game, having to be logged in to use it makes it less useful than something like Ventrilo or Mumble.\(^{10}\)

This supports the conclusion of section 4.4.3 that an integrated voice channel such as that provided in DDO provides only limited benefit, especially for experienced players in long-lived guilds who have evolved sophisticated voice channel management.

### 4.4.7 Reflections on and limitations of research methods

At the start of this discussion section I commented on the likelihood that results from 15 participants in Australia would be generalizable to the broader MMOG-playing population. Though my sample was small, I would argue that the *variety* of experiences reported was itself a finding which will naturally generalize to a larger population. Furthermore, the nature of online cultures is such that my participants were likely to have been sufficiently in touch with

---

global MMOG culture (leaving aside the question of multiple cultures) to have had broadly representative experiences.

With a small study like this, the trade-off against generalizability is richness of data about the everyday experience of users (McCarthy and Wright, 2004). The data presented in this chapter reflect the kind of richness that ethnographic methods can produce.

### 4.5 Conclusion and next steps

MMOG players perform identities and occupy locations in both virtual and physical spaces (Nitsche, 2008). They are real people interacting with other real people, but they are “dressed” as fictional characters in fictional settings. Reality and fiction intersect at the user-interface, and especially in user-to-user conversation, which can expose the offline reality of users and their physical surroundings. Voice, especially, in conveying more information about the person’s offline identity than does text, can destroy whatever barrier users might want to create between physical and virtual worlds. The suitability of different media for communicating in MMOGs depends on a dynamic compromise between the need for people to understand their collaborators and their desire for privacy and a fluid identity. Text may persist as the best option in some MMOG scenarios.

Much of the experience of MMOG users accords with existing CMC theory. For example, in concordance with Lober et al., (2007), users reported that the usefulness of voice relative to text decreases as group size increases. Role-play is not dealt with in existing CMC literature; however one could argue that it represents a kind of “impression management”, which makes role-players’ preference for text compatible with the findings of Carlson et al. (2004). MMOG voice channels seem prone to domination by extroverts, as described in France et al. (2001). The problems that arise through being overheard by third parties to MMOG conversations are similar to those experienced by users of mobile phones, as reported by Faulkner and Culwin (2004). However the unique combination of features that influence MMOG communication requires either refinement of existing communication theory or development of new theory.

Rather than concluding that either voice or text is straightforwardly “better” for this type of virtual world, one must conclude that the different properties of these two modalities make them more or less suited to different situations. The framework best suited to explaining this is arguably that of media synchronicity theory (Dennis et al., 2008), which emphasizes that media differ along multiple dimensions. Voice offers superior feedback and symbol variety –
speakers can convey more concepts more quickly - and this is useful for team coordination and dealing with urgent problems. Text chat affords better parallelism – that is, more conversations can occur at the same time - and this is useful when engaging simultaneously with different people in a complex VW. The reprocessability of text allows people to read back through a conversation, which is an advantage when users are distracted by gameplay activity or have been away-from-keyboard.

Most communicators today have a repertoire of media to choose from. Users who adopt voice retain the option to use text in situations where the latter is better suited, and can fit the medium to the task, such as is discussed by Watson-Manheim and Belanger (2007) and Zigurs and Buckland (1998). However, in agreement with Fulk et al (1990) and Markus (1990), the decision to use a particular medium may often be made by the team rather than the individual.

This study contributed to communication research by elucidating the particular combination of features that comprise communication scenarios in virtual worlds. It contributed to VW research by illustrating the variety and the important influence of the physical world contexts in which VW users find themselves. Finally, the study helped inform the practical design of communication channels for MMOGs, by highlighting problems and some possible solutions, and showing that voice alone is unlikely to adequately support the needs of users.

This chapter is long because many of the themes of the thesis first emerged in study 1. I attempted to examine the influence on user communication of as many different aspects of MMOG usage as possible. However the influence of one important feature of virtual worlds – their spatiality – was difficult to address in study 1, because of limitations in the voice products available at the time. The ‘radio’ configuration used in this study abstracts away users’ virtual locations. To investigate how voice works in 3d virtual space would require a new kind of voice system. Though no spatial voice system was commercially available at this time, I was in contact with researchers who had a functioning prototype that was ready for user-testing. I took the opportunity to investigate this system in my second study, which is described in the next chapter.
Chapter 5

Study 2: Spatially-propagated voice

5.1 Background

Recall from section 4.3.18 that some of the participants in study 1 discussed the possibility of a voice channel that took avatar location into account, analogous to the proximity text channels that were offered in most MMOGs. In such a channel, when someone spoke, their message would only be received by people whose avatars were near the speaker’s avatar. Other participants felt that voice shout be made directional, to give a cue as to which avatars corresponded to the speaker. Participants in study 1 felt that these properties would help solve the problems of channel clutter they had encountered and thereby make it possible for a voice channel to be used by large groups, or even be public, rather than restricted to a pre-defined team.

At the time I began study 2, all existing voice systems in virtual worlds utilized the ‘radio’ metaphor. That is, the communication was organized into channels, and all of the users logged into a given channel received the same utterances. Communication was unaffected by users’ actions and locations in the virtual world.

Some researchers (e.g. Terrano, 2003) had experimented with a spatial configuration that mimicked the transmission of sound in air. According to this metaphor the volume at which an utterance was received decreased with distance from their avatar. My supervisor and I took the opportunity to conduct a trial of one such system which had been created by Boustead et al. (2005).

The system studied was under development at the time of the trial, and has been known by several names. The most evocative of these has been SpatialVoice, and that is how I will refer to it in this thesis. (It is currently marketed by Dolby as Axon\textsuperscript{11}.) This was one of the first

\textsuperscript{11} See https://axon.dolby.com/ for information about Dolby Axon
spatial voice systems developed, and this study was the first reported user-trial of such a system.

Whereas study 1 and previous work showed how the use of voice influenced aspects of virtual worlds such as the use of pseudonymous avatars, study 2 showed how voice could interact with the simulated three-dimensional space of a virtual world.

5.1.1 The SpatialVoice system

SpatialVoice (SV) was designed to provide a realistic sound scene for multiple players communicating in a virtual environment, so that users could converse as though they were speaking to each other in physical space unaided by technology (Boustead et al., 2005). In team-based games, SV did not restrict its spatial mode to team-mates. Members of an enemy team could also hear a player speaking if they were close enough. SV was also stereo-directional, in that utterances appeared to come from the speaker’s virtual location, taking the orientation of the receiver’s avatar into account. This required that SV have real-time knowledge of all of the avatars in a game.

SV also allowed users to switch temporarily to a ‘radio’ broadcast mode. In this mode all members of the user’s team could hear each other equally regardless of location. (This is the propagation rule that most existing voice services for games used.) Radio mode did not mute the transmission of a user’s voice into their avatar’s vicinity.

In the ‘sound in air’ mode, properties of the environment such as the placement and reflectivity of objects were taken into consideration when calculating transmission, in order to make the audio scene as realistic as possible. For this reason, whereas radio systems can be independent of the game being played, SpatialVoice had to have knowledge of the game world’s map.

At the time of this study, SV had been integrated with two popular team-based first person shooter games: *Counterstrike Source* and *Wolfenstein: Enemy Territory*. As noted in chapter 2, FPSs are not virtual worlds, as their spaces are created anew for each game session; they do not persist from session to session and are not inhabited by large numbers of users. However no spatial voice service was yet available in true virtual worlds. As discussed in the previous chapter, a significant portion of MMOG play (especially raids) is similar to the team combat of FPS games. Therefore a study of spatial voice in a FPS would provide a good approximation of the influence that spatial voice might have on virtual worlds.
Therefore I conducted a study of the use of spatial voice by people playing Enemy Territory. ET was a world-war-two themed, networked FPS in which two teams competed during individual game sessions that typically lasted 30 to 60 minutes. ET was played over a local area network with each player seated at their own PC. One player created a server and the others connected to it.

The gameplay in ET was similar in many ways to that found in Counterstrike (Reeves et al., 2009). Upon login the players formed into two teams. Following the theme, one team became the ‘Allies’ and the other, the ‘Axis’. One team was presented with an objective such as blowing up a bridge, which the other team attempted to prevent. The various members of a team had to carry out different roles such as infantryman, engineer and medic. Success depended heavily on coordinating these roles effectively, and therefore good communication was critical.

5.2 Method

5.2.1 Aim

The aim of study 2 was to answer the research question:

RQ2: How does spatially-propagated voice influence social interaction and tactics at the ‘macro’ level of spatiality in multiplayer videogames?

5.2.2 Participants

The participants were a group of employees at a large institution in Melbourne (not my university). Nine participants were recruited through professional contacts to take part in the study. These were technology professionals, but did not work in the fields of virtual worlds, videogames or VoIP. All were males aged between 20 and 40. Participants were not compensated financially in this study.

All participants regularly played videogames for at least 2 hours per week – some for much longer – and many were playing online games. They reported their favourite games to be Counterstrike and Doom and real-time strategy games. Half of the participants had used voice
clients designed for gaming including Roger-Wilco and Ventrilo. This allowed some of them to draw comparisons between SpatialVoice and existing systems.

The participants knew each other well and had been playing a weekly game of Enemy Territory (ET) using their workplace LAN for several months. This session usually lasted around 1.5 hours. Most played using the desktop PCs in their respective offices, and were out of each other’s sight, and in some cases in separate buildings. Their geographical dispersion made mediated communication necessary.

Prior to the study, the participants had used ET’s built-in text-chat facility in order to communicate while playing. However they were not satisfied with it, feeling that typing was unsuited to the coordination of a fast-paced game.

5.2.3 Materials

The SpatialVoice software was installed on each of the participant’s PCs.

Each participant was provided with a standard headset comprising stereo headphones and a microphone.

All participants were given a simple paper diary each in which to record experiences. Play sessions and focus-groups were recorded with a video camera.

The game Enemy Territory was already installed on the participants’ PCs.

5.2.4 Approach

The participants were asked to continue playing their regular weekly game of Enemy Territory, except that during the study period they were to use SpatialVoice to communicate while playing.

The study consisted of two phases, though this was not originally planned. The first phase lasted only a month and was concerned mostly with usability and technical problems in this recently-developed product. These problems proved to be significant, and by the time they were fixed to the point where the participants could freely use SV without problems, the planned trial period was nearly over.
However the participants had begun to enjoy SV and asked if they could keep using it. The SV developers were happy for this to happen. Therefore the study entered a second phase. Phase two, comprising regular use of the now-mature product, lasted two months and culminated in a final focus group discussion.

Most of the data presented here are drawn from this final focus group, as it was there that most of the interesting concepts relating to spatiality emerged.

### 5.2.5 Data Collection

**Phase One:**

With a colleague I carried out observations of individual users: two participants were directly observed and videotaped during each game session. We rotated observation so that each participant was observed at least once. This produced data on problems with setup, and how SV influenced the users’ play and their interactions with each other.

Each participant kept a diary in which they were encouraged to record their experiences of using SV.

All participants joined in a pre-trial group interview a week before the trial commenced. Questions sought participants’ demographic data, their history of game play and their experience with other voice technologies.

At the end of this phase they participated in a focus-group session to discuss SV.

**Phase Two:**

No recording or observation occurred during phase two. Participants continued to use SV for two more months. At the end of this period they participated in another focus group interview.

The timeline in figure 10 illustrates the data-collection process for study 2.
5.2.6 Data Analysis

The data were analysed into a set of themes and presented at a conference (Gibbs et al., 2006). While writing my thesis I re-analyzed the data using NVivo, and refined my coding hierarchy. The coding scheme emerged from the data but was also sensitized by themes emerging from other studies. The resulting hierarchy along with sample data from this study can be inspected in appendix E. The results are presented in section 5.3.

Recall that some of these participants were encountering voice in a virtual space for the first time. Because of this, some of the results refer to the comparison between voice and text, and can be directly compared to study 1 and earlier findings. In the Results section below these findings are described first, under the headings of ‘Comparing voice with text’ and ‘Problems with voice’. Following these are findings that relate to the comparison of spatial and radio propagation. Finally there is a miscellaneous category for participant experiences and opinions that were non-trivial but did not fit the two major high-level categories.

5.3 Findings

Many of the findings in phase one were concerned with technical and usability problems, and were reported to the developers of SV. These are of limited interest within this thesis and I have mostly omitted them, except for some comments under the theme of ‘technical problems’.
The focus-group held at the end of phase two was concerned more with the influence of the spatial made, and it is from here that most of the data for this chapter were obtained.

Because some of the participants had not used voice in games before this trial, some of their comments compare the voice and no-voice conditions. Often in the transcript these comments appear first, before discussion of spatiality arises: I interpret this as indicating that the participants found the difference between voice and text to be the more compelling than the difference between radio and spatial voice. Therefore I present the former topic as a prelude to the topic of spatiality, and because it offers an interesting comparison with study 1.

5.3.1 Comparing voice with text

The participants in study 2 overwhelmingly favoured voice communication over text. They felt that voice had become an integral part of their gaming experience.

Now that we’ve used it for much longer since the last focus group it’s actually become part of the game as well. You actually start to miss it when you’re shouting and someone’s not responding. You’re actually using it a lot more.

ET is definitely more fun with SpatialVoice.

Participants found voice easy and natural to use, compared to text which forced users involved in frantic activity to stop and type. “It’s very easy and natural to do that with SV. You can just react to a situation immediately and call ‘look-out’ or whatever, and you don’t think twice about it.”

Some had, during the trial period, tried playing ET on a text-only public server. They felt the game was lacking when voice was unavailable. “For the first time in a few months I tried to play ET online on the weekend, and I really missed SpatialVoice. It was extremely quiet. [laughter]”

In particular, participants felt that voice made playing more of a social experience. Teamwork became more important. Voice transformed the experience of playing.

Yes I think the nature of the game is completely different with SV and without SV. Like if you play on an online server [not equipped with SV] you have a team, but for all intents and purposes they’re meaningless to you. It’s about your stats and what you’re doing. Whereas with SV it’s a lot more about how the team is performing as opposed to how you as an individual are performing.

Participants felt that voice made the game world seem more immersive. One suggested text should be disabled, as this now distracted from the realism of the experience.
Make it so that SV is the *only* level of communication you’re allowed to use. Make it so that your in-game typed messages don’t work. You just have voice communication. That would add a level of immersion that you couldn’t achieve with other tools.

Using voice, teams engaged in more planning. Members moved as a group and felt more a part of their team’s collective action.

*Did you find that it’s changed your actual game-play?*

Definitely. Before, we never planned anything, we’d just run off. [laughter]

Now pretty much before each game we seem to huddle together and decide, ok you’re going to go ahead, you’re going to be medic, and let’s try to have some sort of a plan. Before it was pretty much everyone on their own.

Essentially the team is running off as a group of say 4 or 5 guys, on a particular map. Now in those 10 seconds from point A to point B, we might have already hatched a plan on SV, and be discussing the plan, use that opportunity while we’re moving as a group. Whereas [on the public server without voice] you’re not.

A plan that failed disastrously was enjoyed as much as one that worked. Participants enjoyed discussing such situations with others while the game was underway.

And then, five seconds later you see your plan totally disintegrate. And then we’re screaming at each other: ‘Where are you going?’” [laughter]

It would be difficult to achieve social interaction of this intensity in a combat game using the medium of text.

### 5.3.2 Problems with voice

The participants’ experience of SpatialVoice was not entirely positive. There were frequent technical difficulties and it was some time into the trial before SV worked reliably. (This was partly due to SV’s being under development.)

*So did you feel it met those expectations, or did it turn out to be different to what you expected?*

Well initially it didn’t work. [laughter] So the first few weeks of the trial we were fixing it essentially. And it would work ok for the next week, then it wouldn’t work, and there’d be network issues, but it’s all fixed now I think. Now it’s good. I guess the initial experience was probably not what we were hoping for, because it had so many teething problems before it got ready.

Participants reported a number of problems, some of which resonated with findings from study 1 and earlier research. Although SV read data from the game map, it was not fully integrated with the game software, so that users needed to take care to use the same user-ids in both systems to keep track of who was speaking. It was not obvious which collaborators had joined the voice channel, so trouble-shooting and testing was necessary, especially at the start of
game sessions. Problems with distorted sound, too-quiet sound, echo, and the Windows firewall were common.

Unlike text, which can be delayed for seconds before being noticed, speech descended quickly into incomprehensibility when CPU or network overload caused brief interruptions.

It was a bit choppy.

If the audio quality is really bad and you can’t understand it, you don’t want to hear it. You almost want it to switch itself off, then switch itself back on when the connection repairs itself.

Technical problems early in the trial mean that in some sessions teams included members who were unable to use voice. As reported in study 1, such teams tended to fragment into voice and text groups who didn’t communicate with each other.

I noticed, when we first played, that some people had SV and some didn’t. We had a class structure even in our own team. You’d only hang around the guys who had SV. [laughter]

This disadvantages people who don’t have SV, if the only means of communication you’ve got is typing, and all the people with SV are ignoring text, no-one’s listening to you.

Another problem that resonated with earlier findings was unequal loudness of voice across different team members. As noted in study 1 this is difficult to control because it depends on factors external to the game software such as microphone placement. Potentially this is more of a problem in a spatial voice system as the volume of an utterance is intended to indicate how far the speaking user’s avatar is in virtual space.

I certainly noticed when we started that the difference between people’s microphone levels was really distinct – some were really quiet and some were really loud.

But over time, technical problems attenuated and SV was fully appropriated by the participants. People stopped typing into the text channel, and soon stopped reading it.

One thing I’ve started noticing is, since I’ve been using SV, I’ve been ignoring the text channel, so if someone types on it, even though it’s right there in front of me, I just don’t seem to see it, whereas before I actually would have.

For me it was infuriating if people didn’t have SV, because you’d be talking to people and expecting them to be able to understand you and they’d just walk off. At the start a few people didn’t have SV, didn’t want to have SV, but by the end of it everyone had SV.

In accord with earlier findings, participants who had used voice in games before noted that the utility of voice depended greatly on whether users knew each other. They were especially disdainful of using voice channels on public game servers.

*Is the fact that you are talking to people what makes it important that you be playing with people you know?*
Yes it’s not necessarily that people are being more serious about it, because most of the time we’re not that serious.

There’s a huge difference between playing SV here and playing it on a public Internet server. You see people playing Counterstrike with voice on, sitting in front of their stereo with Nine Inch Nails blaring out the entire game. There’s all these issues that happen when you go from a community who know each other to a community that just wants something to do, to muck around.

You need a ‘voice ignore’ command on a public server, to ignore people.

You hear people meowing on there [laughter], all sort of idiotic things, five-year olds …

Participants sometimes had trouble knowing who had spoken a particular utterance. In response to this, the developers of SV implemented a visual indicator that appeared above the avatar of a speaking player. Since these participants knew each other well, this was probably less of a problem than in situations such as MMOG pickup groups, or strangers meeting in VWs.

It wasn’t until much later on in our trial when we got the little speech-bubble thing appearing above people’s heads when they’re speaking, so you can tell who is speaking. Because all you could hear was “three people down, off to the right” or whatever.

It made a big difference, the indicator. I only played once or twice with that, but I can see that would help a lot.

Again in accordance with study 1, participants felt that abusive language could be a problem; more so if people didn’t know each other.

The other thing is it’s fine to use SV with people you know, but if you were playing with people you didn’t know … because I get quite abusive, [laughter]. It’s just part of playing a game. But if you were playing against five-year-olds, you wouldn’t want to be talking the way I talk on SV.

Maybe there needs to be a filter. You can kick kids on CounterStrike servers for typing in rude words, so maybe you should get kicked for saying rude words.

One participant noted that the channel’s spatiality meant there would be fewer witnesses to any abuse of the voice channel. “If it’s very spatial and no-one else can hear you getting abused, you’ve got no-one to back you up.”

Some suggested that restricting voice communication to experienced players, or making it a reward for game achievement, might reduce abuse and accidental misuse of the channel.

The other thing for ET would be to only get the walkie-talkie when you get to a certain level of experience in the game. When you have 400 experience points you’re allowed a walkie-talkie.

So you reach level 2 or 3 and suddenly you get a walkie-talkie and can talk to your team.

Participants sometimes could not reconcile a user’s voice with their avatar’s appearance.

I found a bit of a disparity between seeing characters in games and then hearing their voices. “You’re not a large man!” [laughter] It just doesn’t sound right, because these people are in the
midst of battle, they’re being gunned down, and then you hear a calm voice coming through saying “Watch out to your left”. [laughter] And they just got shot three times in the head.

These participants all knew each other well and unlike in MMOGs, there was no opportunity for avatar-tailoring or confusion about offline identity in this game. However they wondered whether in an online game, voice might reveal facts about a user that were hitherto unknown by team-mates.

Something I’m interested in, if they put SV on a public server is: you’re in a clan and suddenly you hear someone speak and find out they’re five years old. It’s like, “I’m sorry, your clan’s over here”. [laughter] They might be great at playing the game but there’d be this whole other real aspect to it where the five-year-olds are roaming the countryside as the Horde, running up to thirty year olds …

The following discussion fragment illustrates the problems that occur, as noted in study 1, when voices in virtual worlds are overheard in the physical world or vice-versa. In this case a participant accidentally broadcast a landline phone call into the SV channel.

_Did you find your voice was going out into the room you’re sitting in?

Well it’s a bit of a problem at work [laughter]

Only if everybody’s not at lunch. If they’re all at lunch, who cares.

At one stage I was playing in a cubicle and two of my colleagues were there, working through lunch, and I was getting these little snide comments every now and then. So I’m going to move next week. [much laughter throughout this topic]

I had headphones on – it’s just the fact that I was screaming down the mic.

Headphones make you louder!

I took a phone call, and I just took my headphones off, and [broadcast the call into the game - laughter drowns out sentence]

I had someone come to my desk and ask me questions, so I paused the game, and I realized I still had my mic on.

Every now and then there’s a sound like someone’s phone going off. Their phone is near their microphone. “Ring ring” [laughter]

We’re playing in a work environment. I guess the majority of people would be playing in a situation where they were just playing games, and wouldn’t have those sorts of interruptions.

(Contrary to this last utterance, study 1 showed that non-work environments are just as prone to problems of unintentional eavesdropping on the game voice channel.)

The findings presented so far were related to the comparison between voice and text. In general these accorded with the findings of study 1. The remainder relate to the comparison between spatial and radio voice, and derive from the focus-group held at the end of phase two.
5.3.3 Spatially-propagated voice

The key phenomenon under observation in this study was the spatial propagation of speech in virtual space: SV was one of the first technologies to implement the sound-in-air metaphor and the participants had not encountered it before. There was some confusion over exactly how spatial propagation worked, as this conversation from a focus-group illustrates:

I’m not sure if I understand how some of the spatial stuff works. If I talk at my normal level, does it get quieter as it goes outwards?

Yes. If someone’s talking to you and you run away from them, they’ll fade off.

So if I shout will it go further? If you shout do you project your voice?

I think it does. I saw someone across the battlefield and said “come over here”, then “COME OVER HERE”, and they turned around.

The SV system allowed users to switch to radio mode via a hotkey. Participants tended to use the familiar radio propagation mode (which they called “walkie-talkie”) more in the early part of the trial; later warming to the spatial mode.

I didn’t notice many people using the walkie-talkie very much.

Yeah that’s one thing I noticed that’s different. When we started using SV I would always be on the walkie-talkie, I found that much more useful than spatial. But as time progressed I’m using walkie-talkie less and using spatial more.

While observing game-play early in the trial I concluded that participants regarded the spatial mode as a restriction over the more familiar radio mode. A radio transmission is heard by the whole team: what advantage would there be to only broadcast to a subset of it? However over time, participants discovered advantages of the spatial mode and used the radio mode less often.

They discovered for example that spatial voice helped to filter out irrelevant chat, reducing the traffic they received on the voice channel.

Do you think spatialization adds anything?

Yes I think it makes a big difference because you know you can talk just to the people around you.

If you’re using Ventrilo and you say “I need some help”, you’ve got three medics in the team and they’re saying “where are you?” Whereas if you’ve got spatial audio and you see the medic, there’s only going to be one medic around who can hear you.

Spatial propagation meant that a call for help was automatically addressed only to participants who were nearby and most likely to be in a position to help. Likewise, warnings such as “two enemies round the corner” and “look out - landmine” were received by people to whom they
were most useful. People hearing an utterance knew it was coming from someone whose location made it relevant.

Directions and commands are usually situationally relevant. Utterances like “go left”, “follow me” or “get the truck” are highly indexical. Awareness of proximity allowed participants to make sense of where these utterances were relevant.

Alternatively, some participants felt that using radio mode – speaking to the whole team at once - afforded better coordination of the team, but that this was offset by the realistic nature of the sound-in-air metaphor.

I think products like Roger-Wilco and Ventrilo are targeted at people who want to achieve higher performance as a team, because for example you can speak when you’re dead, or you can speak only to your team – the enemy can’t hear you. All these aspects that give you an edge over your opposition. Whereas SV is about adding an extra element altogether to the game.

Some felt that SV’s tighter integration with the game relative to third-party products could be a disadvantage to teams who merely wanted to win.

If you’re comparing Ventrilo to SV, and you want to win and that’s all you care about, then I’d be using Ventrilo, because of the ability to keep speaking when I’ve been blown to pieces. With SV I’m out of the game, I’m gone.

In this situation – where a user’s character has been killed in the game – Ventrilo’s independence from the action of the game meant that conversation went on uninterrupted. This added to the realism of the experience but was less conducive to winning.

But other participants felt that spatial voice offered a more natural form of communication, and found its ‘relevance filter’ advantageous for team performance.

You can be with a group of people and you might spot someone there and say “look out, they’re just about to shoot you “, and [they only hear you if] they’re right in front of you. This can help to improve performance. If you’re calling out to everybody on a map “hey look out round the corner” everybody’s going to turn around, when you meant it to be for the person standing right next to you. For me it’s the natural way of using language if you were in that situation.

Yeah, with Ventrilo you normally would give that information, but you’d have to be more descriptive. You’d say “there’s someone round the corner, at the North Guard.” So anyone around the North Guard can go “ah, that’s information for me” and anyone who’s not there will go “oh, “.

This participant felt that the time saved by not having to specify to whom an utterance was intended, and the avoidance of confusion in the channel, outweighed the inability to communicate with the entire team.

Radio mode certainly was useful in situations where the whole team needed to be contacted quickly, and participants exploited their ability to easily access it. “The overall walkie-talkie can also be useful if there’s an ultra-emergency - an override like ‘run in quick’.”
SV gave cues both to the distance and the direction of a speaker. Some participants said that the distance cue was the more useful.

You notice the proximity of voice, and that’s an important aspect of SV, but as to whether someone’s to the left or the right of you, I never really paid much attention to that.

But for me, the proximity – I can hear them therefore they’re close to me – was a lot more important than just whether they’re coming from the left or the right.

This is interesting because one would expect that a cue to direction would be very useful during combat. However the stereo sound scene (in any game using this hardware arrangement) can only be presented relative to the avatar’s, not the player’s, gaze direction. These do not always align, so it is possible that this presentation was not natural enough to be easily comprehended.

### 5.3.4 Multiple parallel conversations

An important result from study 1 was that when many users share a single voice channel, multiple overlapping conversation threads often developed, causing problems with turn-taking and making communication difficult. By contrast, spatial propagation allowed small, non-overlapping conversation groups of two or three people to form and disband easily and fluidly as a team dispersed within virtual space. It was as though a channel was automatically created in real time, based on who was in close proximity, Thus congestion was reduced and multiple conversations enabled, without the need to manually configure or change channels on the fly.

This was especially useful in ET because teams tended to divide into subsets who travelled together to tackle a particular task.

In a lot of the maps in ET there are multiple objectives. So you might have 6 people in your team, and 3 do one thing and 3 do the other. You can communicate specifically with your group, your subset of the team, and you can operate more efficiently in that small group. Whereas if you’re using global communication, I’ve played that on CounterStrike and it can get confusing because people are talking about stuff that’s not relevant to you in your position in the game. To convey a message about where you are or what you want to do can be a lot more cumbersome than having the spatial communication with your team.

Study 1 and prior research found that channel clutter became more problematic as group size increased. Spatial voice could ease this by automatically creating conversation groups at the places where action was occurring.

If you have 20 people speaking over Ventrilo, you’ll get absolute chaos. But if you have pockets of 4 or 5 speaking …
With SV you’ve got a small team of 4 or 5 running around trying to do objectives and talking to each other.

Some of the participants were MMOG players and these speculated about the potential for spatial voice in those virtual worlds. They that felt because of this ability to scale with group size, spatial voice might be useful when raiding.

_So you think voice would be useful in a WoW raid. Do you think spatialization of voice would be useful in those environments? Useful or fun or just add to the game?_

To an extent proximity would have to be there because there are hundreds and hundreds of players on the same server. You don’t want everyone talking over the top.

However they felt that spatial voice might work best in less heavily populated areas of an MMOG, rather than at “inns” and other meeting places, where many people were closely co-located, and hearing too many voices would lead to aural chaos.

There are areas with hundreds of people standing around, you don’t want it to be spatial then because the load on the server would be too much. And you don’t want to see them all chatting to each other in different corners of the area as well. You only want spatial in the open environments like the big towns and things like that.

The participants felt that latency due to network lag would present problems for international teams using voice in MMOGs. Intercontinental ping times might be so long as to render turn-taking impossible.

Another problem is that all the Blizzard servers are in the US, so if the SV server was in the US we’d have a huge amount of latency. Because it’s an international game server, you’re going to have people connecting from the US, China, … and the latency between all of them is going to be a problem. You have someone in the US trying to talk to you and you get it half a second later. Then you respond and they get it half a second later. At least with a game like ET the majority of people are connecting locally, so the worst you’ll get is 60 or 100 milliseconds peak, and that’s not too bad. But once you get 400 or 500 ms or up to 1000, game play is fine at those latencies, but audio wouldn’t come out very well.

This last comment illustrates that conversation, more so than game-play, is sensitive to delay.

### 5.3.5 Overhearing the enemy

An interesting aspect of spatially propagated voice was the ability it gave users to hear the voices of members of the opposing team. This was in contrast to existing radio-style products which invariably restricted communication to a user’s own team-mates. In radio voice, enemy players are seen but not heard. This new ability to listen to and even converse with the opposition was exploited by participants in order to devise new game tactics.
For example, some participants enjoyed taunting members of the opposing team: “It forces a level of intimacy too. You can say things to someone right next to you, an opponent who you’ve just killed. Taunting.[laughter]”

Not only did SV make the enemy available for interaction, but the fact that speech was localized made ”trash talk” with the enemy more appropriate, as the entire playing group would not be exposed to it. “You don’t want to sit there on Roger-Wilco and taunt them to everybody, but you’ll do it if it’s just you and them, as a conversation.”

In the Enemy Territory game, injured avatars sometimes lay incapacitated on the ground for a time before ‘dying’. Participants who found themselves in this situation reported making use of the forced delay to taunt the opposition players who had just beaten them, thus distracting their opponents’ time, and to report the enemy positions thus discovered to their comrades. One subject said he encouraged opposition players to take the time to “finish him off”; reasoning that this wasted the enemies’ time and diverted them from more important objectives. “I find the best thing is being able to heckle someone to kill you completely. You pull someone’s attention away from looking around for other enemies.”

In ET there is also a short period while the game starts during which players are free to roam the virtual space but are unable to interact with it. Some participants discovered that during this period it was possible to lurk near the opposition team’s starting area, and listen to their conversations in the hope of discovering their plans. “I remember we often used to send one person out during the warm-up phase to listen to what the other team was planning.”

They also found that at any stage in the game it was possible to overhear enemy players discussing tactics, and were able to use this information to their advantage. Spying and eavesdropping thus became a new strategy that was not possible in other voice systems. This added a new aspect to gameplay, but a new danger as well: the voice channel now had to be used with care.

Sometimes you’re in a position where you can hear the enemy. They might be hiding around a corner and you hear them say “throw an air-strike in” or they’re talking about the tactic that they’re going to use. So you can really take advantage of the knowledge of what they’re going to do. This wouldn’t be there with Roger-Wilco or other systems. So it definitely can be a disadvantage to be using SV, but I think it does add to the game-play.

Conversely, participants who were aware they could be overheard by the opposing team could attempt not only to avoid espionage, but to actively mislead the opposition team.

During one game, two participants tried to create the impression of greater numbers by running to different points in a building and calling out in different voices.
We were guarding a room where the opposition has to run in and storm it to get something out of it. We came up with this plan that we would speak quite loudly and put on a whole heap of different voices so that it sounds like the room’s packed with opposition.

Not all participants approved of the ability of opposing teams to hear each other’s utterances. One of them felt that the danger of being spied upon represented an impediment to the performance of teams using SV. A team using radio voice might have an advantage over a team using spatial voice.

I think products like Roger-Wilco and Ventrilo were targeted at people who want to achieve higher performance as a team, because for example you can speak when you’re dead, or you can speak only to your team – the enemy can’t hear you. All these aspects that give you an edge over your opposition.

In this exchange, participants suggested an alternative configuration for the “hear the enemy” feature: transmitting garbled speech, so that the presence of the opposition was indicated but not the meaning of their utterances.

Another idea might be: even if you can hear your opponents, you can’t understand what they’re saying. So it comes across like in a foreign language or something.

Then you’re not just doing tactics, you’ve got that spatial awareness.

And you do know there’s an opposition there, because he’s talking another language.

Nine times out of ten you’re just going to go and kill the guy anyway. You don’t really care what he’s saying.

You might be able to get some information – torture him! [laughter]

However most felt that the ability to overhear the enemy was a valuable addition to the game environment, enhancing their fun and the game world’s realism in spite of the opportunity to be spied upon.

What about a competition mode where the vicinity chat worked within your team but wasn’t broadcast?

I think it would detract. If you’re going to have it positional, it should be for everyone.

I think that’s part of the fun of it. I like the fact that you can hear opponents. That’s one of the things I was really looking forward to in the trial – it wasn’t just an intra-team communication. To me, that’s done, it’s pretty tame. But to have this – I think the real benefit is that it does immerse you in the environment more. I like that aspect where you’re lying on the ground and taunting opponents.

In this case both team-mates and enemy players were offline friends, and the desire to engage socially with them overrode concerns about success in competition.
5.3.6 Likelihood to adopt

Participants had mixed feelings about whether they or others would be likely to pay for SV as a distinct service. On the one hand, some felt that SV made such a dramatic difference to gameplay that once exposed to it, they would pay to keep using it.

If I had a 30 day trial using SV and then went back to the game without SV, I’d miss it so much that I would say, yes I definitely want to have the SV functionality because I enjoyed it so much.

I’ve got to a point where I don’t enjoy playing ET any more without SV. So I would pay for it. ET is a good game but it’s so much better with SV.

But one participant felt that that the advantages of spatial voice would best be enjoyed by advanced game players who knew each other well.

I would pay to play with this group of players, but not to play with people who hadn’t been exposed to it before. I see the advantage of SV within this particular group. [Playing with non-experts] would seem like you’re getting pretty much the same thing as Ventrilo.

Spatial voice without the expertise to use it effectively would be no more advantageous than radio voice.

Others felt that rather than pay for SV as a standalone product, people might pay a premium for a game that incorporated SV.

Some participants were concerned that, while spatial voice presented several advantages, teams using radio mode might enjoy an advantage over teams using spatial mode. They felt that rather than trying to pit spatial against radio teams, teams might be willing to compete in a spatial-only environment. All teams would equally enjoy the filter for relevance, but at the same time be equally vulnerable to spying.

5.4 Discussion

Early implementations of voice channels in virtual environments worked in an “all to all” mode – what I have called the ‘radio metaphor’ - such that when one user spoke, all of their collaborators could hear, regardless of their current location or activity in the virtual environment. The voice systems examined in study 1 worked in this way. Such a channel is relatively straightforward to implement, as the digitized utterances need merely be delivered to multiple Internet addresses, without the need to individually process the characteristics of each transmission.
However while it is relatively simple for a virtual environment to implement a radio mode voice channel, to do so does not take into consideration the defining feature of a virtual environment: its simulation of space. In most respects virtual worlds are designed to give the user the perception that they are located within a three-dimensional environment. For example objects are visually rendered to reflect their distance and direction from the user’s avatar, and can be occluded by an intervening object. Why not then render sounds in a similarly realistic, spatial manner? This study revealed the advantages and disadvantages that can accrue by doing this.

The Spatial Voice channel took into consideration: the structure of the virtual environment, the locations and surface materials of the virtual objects within it, and the relative positions of users’ avatars, when it was transmitting speech from user to user. That is, volumes were attenuated as the distance from speaker’s avatar to receiver’s avatar increased, and according to the presence, size, shape and surface of objects between or around the avatars. A stereo positioning of the utterance at the receiving end could also be calculated, so that the utterance seemed to come from the direction of the sender’s avatar relative to the direction the receiver’s avatars was facing.

This approach to configuring a voice channel is unique to virtual environments, as it depends on the simulation of a space in which the users are located. Therefore this approach is not addressed in prior research on mediated communication. The system examined in this study was one of the first spatial systems to be developed, and this was the first reported study of the use of a spatial voice system.

The aim of study 2 was to answer the research question:

RQ2: How does spatially-propagated voice influence social interaction and tactics at the ‘macro’ level of spatiality in multiplayer videogames?

The extended period of use lasting several months, combined with a UX approach to data collection and analysis, allowed me to present a rich set of findings on the influence of spatial voice on the experience of using a virtual environment, and especially certain subtle benefits that were only apparent after extended use. The study environment enabled a clear comparison between the radio and spatial designs, in a game in which a team’s shared negotiation of space was of paramount importance. This case provided valuable data for the thesis and acted as a bridge between case one (an MMOG with radio voice) and cases three and four (a social VW with spatial voice).
The spatial mode represented a restriction of sorts over the more common and more easily implemented radio metaphor. Because of this restriction, my expectation was that users would initially prefer radio mode, and this was borne out. But over time participants discovered advantages to spatial voice that were not obvious on first exposure.

5.4.1 Constraining communication to make it more intelligible

While the spatial mode seemed at first to be a restriction of the more common radio mode, the constraint in fact made some communication more intelligible to users. In effect, SV added extra socially relevant information to the voice channel. Users received not only the spoken message, but a sense of how close the speaker was, and in which direction, making it easier to associate a voice with an avatar. In team games where rapid negotiation of space is key to success, information about other players’ positions is a valuable resource. Thus what at first appeared to be a constraint on communication proved to be a source of extra information about the state of the game world and its contents.

Spatial voice acted as a ‘filter for relevance’, in that only users close to a speaker, and therefore able to interact immediately with the action in the speaker’s locale, could hear the speaker’s vocalizations. From the point of view of any given player, knowing that they were able to hear game-related speech such as warnings, cries for help, shouts of surprise or commands, meant that significant action was occurring nearby, and that there was a good chance that these utterances was directed at the player, or at least that the player should respond to it. Compare this to the situations encountered by the MMOG users in study 1, where channel clutter made it impossible to know who was speaking, nor to what or to whom they were referring, nor whether their utterances were relevant to the listener.

It is interesting to compare the constraints built into SpatialVoice to the suggestion that systems for computer-supported collaborative work should be “socially translucent” (Erickson and Kellog, 2000). The social translucence principle, reviewed in chapter 2, proposes that systems should deliver not only linguistic information but socially salient information about the speaker, such as who they are and where they are situated. Socially translucent systems are not transparent; rather they are constrained. When participants share an awareness of the “physics” or transmission rules that govern a communication system, they can use this awareness as a resource for structuring their social interactions. Although SV constrained communication, in doing so it provided its users with an interpretive resource for making sense of what each other were saying.
SpatialVoice was one of the first projects to explore possibilities for voice channel configuration beyond the simple radio mode. Voice systems that emulate two-way radio are disconnected from the spatiality of virtual environments, and by and large from the events that occur in them. Integrating voice into the space increased the feeling of immersion, and made achieving and maintaining communication between team members a significant part of team strategy, mimicking what often must occur in physical space. However the difficulty of attaining a consistent volume from user to user probably lessened the sense of realism.

5.4.2 The desire for realism versus the desire to win

It would appear that spatial propagation of voice adds to the realism of a multi-user virtual environment. However it is interesting that there was tension among the participants over whether such realism was always desirable. Some perceived a trade-off between realism and utility. There were times in the game when it may have been more advantageous to simply hear all of one’s team-mates, regardless of where they are in virtual space. Not all relevant action is local in a combat game: for example an attack on a team’s home base might require all roaming members to return and defend it. A purely spatial voice product would render impossible the non-spatially-restricted communication required to immediately gain the attention of all team members in an emergency. (This product was not thus afflicted as it allowed users to switch to radio broadcast when required.) The participants in study 2 felt that if a team using spatial voice was not competitive against a team using radio, then people would not choose the spatial product, even if it provided a greater sense of immersion in the space.

This is reminiscent of Golub’s (2010) argument against the necessity for visual realism in massively multiplayer online games. Golub was concerned with the software add-ons used by hard-core raiding guilds in World of Warcraft which decompose the visual representation of the virtual world into charts that visualize the game’s underlying statistics. Users of these add-ons effectively give up visual realism in favour of a display that makes the raid more likely to succeed. Golub argued that success in the game - the shared project - was more motivating for many players than the sense of immersion or presence in a virtual environment.

Perhaps then it is less important to provide virtual world users with a voice channel designed to immerse them in three-dimensional space than it is to bind them together through successful competition against other teams.
5.4.3 The desire to socialize

My observations of team interaction in study 2 suggest another interpretation as well. Perhaps the experience that an all-to-all radio voice channel simulates is not so much the use of radio, but the experience of playing together in the same physical space.

Recall Taylor and Witkowski’s (2010) description of the intense sociality that occurs in face-to-face interaction at LAN parties, and Aarssand and Aronsson’s (2008) description of the stream of speech and “response cries” that mediates play among children co-located around game consoles. When people play together in the same physical space, they remain in the same conversation space regardless of where their avatars go or whether their avatars live or die. The conversation continues when the game ends, and even when the PC or console is turned off. Using a voice system like Ventrilo, which is completely independent of the game and what happens in it, provides a communication experience somewhat like co-located play. Golub noted that:

> Vent is important not only for giving instructions during raid, but also as a place where the guild socializes, jokes, and chats. The PA vent server is thus a second virtual location, often used in conjunction with but separate from the World of Warcraft game client. (2010)

Supporting Golub, some of the participants in study 1 described using Ventrilo in ways that were independent of in-game events, such as joining the voice channel before logging into the game in order to chat casually or even to decide which game is going to be played, and continuing to chat when their character died or the game was over. This also supports Golub’s claim that commitment to the game, and in particular to achievement and to one’s team, leads gamers to engage in knowledge exchange outside of the game, and that it is the collective project, not simulated qualia, that causes people to be “immersed”.

It is interesting to note then that, since this study into a very early implementation of spatial voice was conducted, the radio metaphor has continued to dominate voice products designed for players of MMOG and FPS games. Study 2 suggests a number of reasons why this could be the case:

- radio voice makes teams more competitive (in most games), and users find this more important than the advantages of spatial voice,
- users perceive radio voice to be a simulation of the experience of co-located LAN game play, and enjoy this,
• incorporating game-world maps into the voice system is too expensive given the number of different games and maps.

On the other hand, as we shall see in the next two chapters, the spatial metaphor was chosen when voice was implemented in the non-game “social” world Second Life. Note that none of the reasons of the last paragraph apply in a non-game world, while it is critical to deal with channel clutter in this large virtual space.

5.4.4 Limitations of this study

Of the four studies presented in this thesis, this was the most limited in its scope. There were relatively fewer participants, they were using a simpler kind of virtual environment, and they all knew each other, preventing the study from examining what study 1 suggested was a key condition for voice acceptance.

This study used an FPS, which was neither massively-multiplayer nor persistent. At the time of this study, spatial voice for true virtual worlds was still a year away.

However as noted above, important parts of MMOG play are similar to FPS play, and study 2 yielded interesting insights into the user experience of a spatial voice channel, serving as a comparison with study 1, and a stepping stone for studies three and four. Furthermore many of the participant’s comments on voice allowed a comparison with the results of study 1.

5.5 Conclusion and next steps

Within limits imposed by technical concerns, spatial propagation of voice made the virtual space more immersive, enabled new tactics, and ameliorated problems with channel clutter discovered in study 1 and earlier work. However features such as overhearing and being overheard by opposing teams, and losing contact with one’s own team, while enhancing the realism of the experience of virtual space, could be detrimental in a combat situation.

Studies one and two were conducted within competitive game environments. Game worlds represent only one of the two major types of virtual world: The grounded theory principle of theoretical sampling suggested that at this stage of my program I should begin to examine the other type: non-game, ‘open-ended’ or ‘social’ virtual worlds.
As noted in chapters one and two, social worlds by definition are not structured around competitive game-play. In particular the team combat scenarios that were important in studies one and two do not feature in social worlds. Since this scenario presented a compelling use-case for voice, it was likely that a virtual world that was absent this scenario would offer a different perspective on the value of voice.

At this stage in my work, the leading social world Second Life was receiving significant attention from researchers as well as the popular press. Coincidentally at this time (2007) Second Life’s vendor announced that they intended to introduce a voice channel. This seemed like an ideal opportunity to observe the introduction of a voice channel into a virtual world and one that was significantly different to the game worlds of the earlier studies.

The type of communication examined in study 2 can be thought of as occurring on the ‘macro’ spatial scale in which teams negotiate large virtual spaces. By contrast, study 3 examined an interaction at the ‘micro’ scale of collaboration around virtual objects. Study 4 examined the general influence of voice upon the user experience of Second Life.
Chapter 6

Study 3: Communication during collaboration around objects

6.1 Background

Studies one and two demonstrated the value of voice communication for sociability and real-time coordination of groups in virtual space, highlighting at the same time problems such as channel clutter, diminished support for role-play, and a reluctance to communicate with strangers.

Study 2 showed that when voice in a virtual environment was designed to propagate spatially, it could enhance immersion, reduce channel congestion, and afford new game-play strategies that both emphasized and exploited the spatiality of the environment.

These studies were concerned with communication within the ‘macro’ spatiality of the distribution of avatars in a virtual world. In such a scenario, the fundamental problem that a communication system must tackle is that users’ avatars are dispersed and moving in a large space, so that it is difficult for users to keep track of each other visually and to coordinate their joint activity. Voice helped to solve this, though it introduced problems of its own.

On the other hand, communication sometimes needs to occur within the ‘micro’ spatiality of collaboration around objects. In the physical world, people gather around artefacts in order to perform work: they might be repairing a machine, designing a house or discussing a map. Here the problem is not keeping track of the location of one’s collaborators – they are clearly in view – but in maintaining awareness of their actions and where their foci of attention are. Collaborative work requires “the up-to-the-moment understanding of another person’s interaction with the shared workspace” (Gutwin and Greeberg, 2002, p. 417).

In particular, collaborators in work around objects must be able to understand, at a fine level of detail, to which object or location their collaborators are referring when they or use gestures such as pointing or expressions such as “this” and “there”. Collaborators co-located in physical
space can acquire information about each other’s activities and the state of their field of work with relative ease (Schmidt, 2002). People utilize the hands, arms, face, eyes and body to deliver non-linguistic signals including iconic, indexical, and demonstrative gestures (Clark, 1996).

At first glance one might expect that virtual environments should support a similar kind of deictic reference, since users are embodied as avatars which can move and gesture. However, research conducted in early CVEs showed that avatar-mediated reference to objects in virtual spaces often failed (Bowers et al., 1996; Hindmarsh et al., 1998; see chapter 2 section 2 for details). These early studies discovered significant problems with turn-taking and dereferencing of deixis.

The appearance of *Second Life* a decade after this early research into collaboration in CVEs offered an opportunity to study the problem in a modern virtual world which differed in being massively-multiuser and Internet-based. *Second Life* represented an advance in graphical realism compared to earlier experimental CVEs. Uniquely among modern VWs, *Second Life* allows users to collaboratively create and modify objects in the virtual space. It also offers an inbuilt voice channel which is spatial and directional. SL has a large user base and a shared culture relating to building objects. This provided an opportunity to compare current practice with observations from older research.

The purpose of my third study was to examine whether deictic reference remained a problem in this modern voice-equipped virtual world, and if so, whether and how people worked around it. There were two *prima facie* reasons to suspect that the problems uncovered by an earlier generation of research, may have attenuated, or at least evolved:

1. 3d technologies had matured – featuring advanced graphics and avatars etc, and
2. more people were familiar with 3d technologies – familiarity with the 3d UI of videogames might lead to more skilful collaboration in a VW.

I conducted the study at the Palo Alto Research Center, where I had an opportunity to work during my candidature. My data-gathering took place during July and August 2008, with the initial analysis occurring during September and October. (I re-analyzed the data later while writing this thesis.)

I gathered data in several ways. To facilitate comparison with earlier research I conducted a lab study inspired by the approach of Hindmarsh et al. (1998), using also some of the methods of Kraut et al. (2002). I asked groups of two or three participants to collaborate on building tasks, and recorded their screen video and conversation for later analysis. I discussed with
these participants the problems they faced and how they solved them, and asked for their thoughts on SL’s user interface. Using recordings I later analysed participants’ deictic verbal references and their use of Second Life’s ‘virtual camera’, an interesting UI feature that I hypothesized would influence collaboration. Finally I discussed themes that arose with other SL users, on the user forum and at in-world events such as building classes.

6.1.1 Prior research

In chapter 2 I reviewed research that emphasized the difficulty that users of CVEs face when trying to communicate about location, a problem which has probably hampered the technology’s successful adoption (Gutwin and Greenberg, 2002).

Hindmarsh and colleagues (2000) used conversation analysis to study groups of users cooperating on joint tasks in an early virtual environment. They found that users struggled to achieve common reference, even when their avatars could ‘point’. Users could not always see both the pointing arm and the referent, due to the limited gestural abilities of avatars and the narrow horizontal field of view of desktop systems. Users repaired failed attempts at deixis through a constant stream of talk through their voice channel.

Moore and colleagues (2007) analyzed avatar- and text-mediated interactions between MMOG players, discovering similar problems despite the greatly increased graphical realism of the more modern systems. They found that avatars did not provide awareness of their owners’ actions, that avatar eye-gaze did not correspond with what users could see, and that gestures were difficult to perform.

Significant progress has been made in making avatars more realistic, usually by improving the fidelity of pointing (e.g., Wong and Gutwin, 2010) or eye-gaze (e.g., Murray et al., 2009). However this approach is not only a significant technical challenge but a usability challenge as well, as these new degrees of freedom must somehow be controlled by the user. Publicly available virtual worlds have not adopted these technologies and continue to rely on standard input devices such as keyboards and mice.

Research in shared-video systems has also shed light on object-focussed collaboration-at-a-distance, though avatars were not used and the choice of vistas was usually limited to ‘scene’ or ‘head-mounted’ cameras. Kraut et al. (2002 and related work) had helper-worker pairs complete a screen-based jigsaw-puzzle. In this arrangement the worker manipulated objects while the helper could offer only verbal direction. Sharing the scene view, but not the head-
mounted views, improved performance, especially when the task was complex and the objects
difficult to describe verbally. Goebbels et al. (2003) had pairs collaboratively manipulate a
virtual object with the assistance of haptic control and video-conferencing, finding that users
spent more time looking at the object than each other, except while resolving
misunderstandings, and that voice quality was more critical than video.

Ranjan et al. (2007) studied methods for automatically controlling a moving camera in shared
video 'helper-worker' scenarios, getting best results when the camera (ie, the helper's
viewpoint) followed the worker's hand rather than worker's viewpoint or other camera angles.
Fussell et al. (2003) found that users do not look at each other so much as look at each other’s
hands and the objects being manipulated. All of this work suggested that it is more useful to
visualize the field of work and a collaborators’ focus of attention than to visualize
collaborators’ bodies.

6.1.2 Second Life’s virtual camera

Second Life implements a feature whereby users can detach their camera – the location in
space from which they are viewing the scene – from their avatar, and move it around the local
area. Hindmarsh et al. (1998) described a similar approach which they called a “virtual
camera”, and I will use that term to describe the Second Life feature.

Because of the virtual camera, at any given moment an SL user is in one of two viewing
modes: they are either looking from their avatar’s position (the default), or they have detached
their camera and moved it elsewhere. I called these modes ‘in avatar’ and ‘in camera’ to
emphasize that subjectively the user may perceive these to be two different locations. ‘In-
camera’ also means ‘secret’, and this is apt because there is no way for an SL user to know
where other users have moved their virtual cameras, nor even to know which camera mode
someone else is using.

These two modes are illustrated in figure 11, which shows the simultaneous vistas of two
participants who are collaborating as part of study 3. The user on the left is ‘in-avatar’, though
zoomed out a little, and looking from a position slightly to the rear and above their avatar’s
head. The user on the right is ‘in-camera’ and has moved his camera above the house to look
down on it. His avatar is out of view. Notice that the prim he is currently editing – one half of
the roof – is highlighted for this user, but not for his collaborator on the left.
Users move from in-avatar to in-camera mode by holding Alt and panning their mouse around an object of interest. The camera centres on that object, and changes location and orientation as the user pans the mouse. SL users commonly call this “alt-zooming”. Users can return their camera to their avatar at any time by hitting Escape.

### 6.1.3 Building in Second Life

*Second Life* allows its users to create the contents of the virtual world. This is a popular activity in SL and users build a wide range of projects including houses, shops and vehicles. Objects are assembled out of simple shapes called ‘prims’ for ‘primitive shapes’, which can be created using the SL user interface. Figure 12 shows an SL user creating and assembling a group of prims.

Limited pointing is possible in SL. When a user is editing a prim, their avatar's arm reaches toward the prim and a dotted line (called the 'selection beam') connects arm to object. Hindmarsh et al. (1998) provided a similar feature, their rationale being that moving an object at a distance represents projection beyond the avatar. In SL this line provides a rough indication of which object another user is editing; however if object and avatar are sufficiently far apart or there are many objects grouped together it is difficult for others to tell what has been selected. The user who is editing also sees a highlight on the object, but this is not visible to others.
6.1.4 Designing the study

The purpose of this study was to examine how groups communicate while collaborating around objects in a modern voice-enabled virtual world. There were a number of ways in which this question might have been addressed. My starting point in designing the study was to repeat the procedure of Hindmarsh et al. (1998). However there were reasons to consider varying this:

1. Hindmarsh had pairs of participants log into a CVE and discuss the placement of furniture in a virtual room. Since Second Life allowed manipulation of objects I felt that moving them to assemble a larger object would be a more thorough test of collaboration.

2. Unlike older experimental CVEs, Second Life had a large user community and significant user culture. This provided an opportunity to refine the lab findings by discussing them with the SL community.

To address point one I designed a protocol based on Hindmarsh et al.’s procedure but borrowing methods from the shared-video experiments of researchers such as Kraut et al. (2002). This work involved pairs of people, separated by a wall but connected by a shared audio-visual channel, who worked together on tasks such as making models or solving jigsaw
puzzles. I adapted the puzzle idea (converting it to 3d) and borrowed Kraut et al.’s ‘worker-helper’ arrangement, whereby one participant manipulated objects under the direction of the other participant.

To address point two I decided the research should be partly ethnographic. A culture of building had developed which was propagated in building classes, interest groups and forums, and through small businesses that developed large ‘builds’ for paying customers. I reasoned that by interrogating expert SL users I would be able to identify the problems and workarounds that modern VW users had discovered.

However while online ethnography had emerged as a popular method for researching virtual worlds, this approach alone would not have worked in this study. Partly this is because it would be rare to chance upon instances of collaborative building in action while wandering about the virtual world. But also, the collaboration problems discovered earlier stem in part from the way the 3d scene is presented on users' screens. To understand these problems I needed to be able to see what users were seeing, and discuss this with them. The only feasible way to do this was to be co-located with the study participants, which doesn't happen in online ethnography. I considered observing SL users in their homes – which would at least have been naturalistic - but to observe a collaborating group I would have needed to place a researcher (or at least a video camera) in each participant’s home, which was impractical.

For these reasons it was decided that observation of building should take place in a lab, where I could observe participants in person, talk with them easily, and record their performance for later analysis.

Being physically co-present with the participants risked lessening the realism of the setting and thereby the generalizability of the findings. However it offered the advantages that participants could ‘think aloud’ while working, discuss problems as they arose, and have questions answered immediately if they didn't understand the task or became lost.

Given this decision, the next question was how to place the members of a group relative to each other in the lab. To fully replicate online collaboration would require them to sit in separate rooms. However this would reintroduce the problem that one researcher could not simultaneously observe all participants.

A related consideration was whether participants should communicate by voice or text, or should be allowed to choose modality. There were several concerns here. First, the designs of study 1 and 2 had facilitated a comparison of modalities. I was comparing voice and text also in study 4 (which was under way), and I felt this reduced the necessity for a comparison in
study 3. In the earlier studies of Hindmarsh and others, participants had used voice. I felt that to ask participants to use text in collaboration of this kind would impose a significant usability burden on them. So I chose voice as the (only) modality for this study.

At the time of the study voice had been available in SL for a year, but it was controversial and had been rejected by many long-term users. For example, people staging events in SL were expected to specify whether the event would use voice or not, so that others could choose whether to attend. I feared that advertising a study that required using SL’s voice system would put some participants off joining. Using SL voice would also have made the trials vulnerable to the vagaries of this system, which was prone to downtime. With some participants travelling to the lab to participate, I didn’t want to risk having to cancel sessions. The technical problems of SL’s voice system are important and are discussed in detail in the next chapter, but this study was not intended to discover problems of that kind. There simply needed to be spoken communication that worked.

To solve both problems at once - the visibility of entire groups to the researcher, and the need for shared voice that worked reliably - I simply assigned group members to different PCs in the same lab. They were arranged so that I could see all their screens, and speak with them easily. They could hear each other speak without having to use VoIP. They could not see each other, as workstations were angled away from each other. I considered placing a room divider between users, but this proved to be unnecessary: I asked participants not to turn to face each other, and only rarely had to remind them of this constraint. I checked with participants who were experienced in SL, and they felt that this arrangement was realistic enough. I allowed participants to use Second Life’s text channel as well if they wished, but few chose to.

Another concern was how to record the sessions. User trials in labs are typically videotaped. However there was a reason not to use cameras in this study. Many SL users, in particular the experts I wanted to attract, had a culture of pseudonymity and preferred never to connect their online and offline identities. This preference extended to a proscription against using real names in SL-related correspondence, and an expectation that others would follow suit. I felt that if I insisted that participants be videotaped or photographed, expert SL users would refuse to take part.

In any case, footage of the participants’ physical movements would have been of little use, as the action was all on screen. I made a simple rule: no cameras would be used in this study.

The screen output on the other hand was very useful, as it included not only what the participants could see, but (when they were ‘in avatar’) their avatar gestures and movements. Matched up with audio of their speech, a screencam recording allowed me later to replay the
trial with high fidelity and analyse the participants’ interactions. Therefore I had participants wear microphones (even though they were not using VoIP), and using the Fraps screencam software I recorded their screen output and speech.

To allow for participants who might not have wanted a recording of their voice to be associated with their regular avatar, I prepared several test SL accounts that participants could use if they wished. Alternatively participants could use their own avatar if they found familiar embodiment to be more comfortable. Most chose to use the avatars I provided.

6.2 Method

6.2.1 Aim

The aim of this study was to answer the research question:

RQ3: How do groups communicate using voice while collaborating around objects at the ‘micro’ level of spatiality in a virtual world?

6.2.2 Participants

I recruited participants from two sources:

- staff and interns at my research institution
- regular Second Life users who were able to travel to the lab

The institutional participants were not experienced with Second Life, though all were interested in technology. Most had played videogames or used 3d editing tools. They were not compensated for participating in the study.

I recruited the SL users through advertisements on Craigslist and the SL user forum. My lab was located in the San Francisco Bay area, a conurbation which includes Stanford University and boasts a high proportion of technology users. Participants received $20 to cover travel costs. Some presented as groups, while others were singles and had to be matched up.

The participants in the lab trials ranged in age from 25 to 50. Half were male. Some knew each other before undertaking the task, while others were meeting for the first time. I chose
participants with a broad range of SL experience, reasoning that while expert users would demonstrate cutting-edge practice, it was important also to understand the novice experience.

I conducted ethnography by discussing themes as they arose with users whom I met in the SL virtual world. (These users were not the same ones that participated in the lab trials.) I found experienced builders by attending events in SL such as building classes and competitions.

6.2.3 Design of trials

The lab participants were formed into groups of two or three, each of which performed a session consisting of two collaborative building tasks.

The first task adapted the 'furniture world' experiment of Hindmarsh et al. (1998), borrowing the 'jigsaw puzzle' task of Kraut et al. (2002), though in 3d. I provided the group with a set of virtual objects including four walls, two halves of a roof, a flag, flagpole, chimney and two gables, placed randomly within the local scene. The group had to assemble these into a house.

Some of the objects, such as the flagpole, were unique and could be referred to by name. Others, such as the walls, were indistinguishable, requiring spatial reference to identify them. I gave each group member a picture of how the house should look when complete. This picture is shown here as figure 13.

![Figure 13: 'House' task screenshot provided to participants](image-url)
The 'House Task' was unstructured, allowing groups to discover their preferred method of collaboration, and naturalistic in that houses are a popular building project in SL. However the house being constructed was relatively simple. The dilemma in choosing the level of complexity was that a complex assembly would have been too hard for novices and a simple one trivial for experts. A possible solution to this would have been to provide a range of tasks tailored to different levels of expertise. However this would have prevented comparisons across the whole cohort.

I considered providing a task more similar to remote repair, as this was a scenario that motivated much work on remote collaboration (e.g. Kraut et al., 2002; p31). A complex machine such as a model engine with moving parts was beyond my building skills. I attempted to buy one in Second Life, and though I found some candidate models, these merely represented the appearance of machines and did not consist of realistic parts. It is possible that the physics simulation in SL is inadequate to represent a working machine. For example, SL allows objects to pass through other objects, obviating the main problem in machine assembly. This lack of collision detection means that object movement has to be carefully programmed. For example screws and nut-bolt assemblies ought only to move in certain ways, and whereas the behaviour of physical versions of these is dictated by their shape, this cannot work without collision detection. SL also makes it difficult to clearly define when a part is in a desired position. For gross movement this is not usually a problem but for simulated repair it must be clearly definable whether a part is in place or not. For these reasons, I stuck with house assembly as a task, and did not attempt to define a completion state or measure completion times.

In pilot studies of the House task I found that some groups chose to collaborate with minimal communication. For this reason I provided a second task which used the helper-worker arrangement of Kraut et al. (2002) to force closer collaboration among the group members. In the 'Garden’ task, one group member was designated the Helper, while the other(s) were Workers. Only the helper received a screen shot (figure 14) showing the house now surrounded by extra objects including garden furniture and a fence.

Helpers could not use the building tools, but instead directed their workers to build this scene. My intention was that this would require frequent communication about location and orientation involving deictic references, such as “you need to place the fence here” or “the chair goes between the table and the house”. I hoped this would surface problems of reference, if they existed, and cause groups to attempt to find solutions to them.
6.2.4 Materials

Participants in the lab trials were given standard PCs with Second Life installed, and a headset with microphone. The PCs were equipped with Fraps screencam software which recorded both the screen output from Second Life and the participants’ voices.

For the ethnographic work no special materials were used.

6.2.5 Approach

I conducted a total of 10 lab sessions. Each participant took part in only one session.

A complete session was designed to last an hour and a half, although some enthusiastic groups worked over time to improve their builds, which I allowed. Most of the trial groups were pairs of participants; however I organized some groups of three to check for differences. Because some groups had three members I have used the word ‘group’ rather than ‘pair’.
After logging into *Second Life* and spending a few minutes familiarizing themselves with each other and the lab setup, the group members brought their avatars to a designated spot in SL. Here I produced the building materials from my inventory, started Fraps recording on each participant’s PC, and asked the group to begin the House task. After this task was complete, I designated one group member as the Helper and asked the group to begin the Garden task. I stayed in the lab with the participants to observe them and answer questions where needed.

After each group had completed both tasks I conducted a focus-group discussion with the participants. I asked them whether they had encountered any problems, what they thought of the SL user-interface, their suggestions for improvements and whether the UI could be enhanced to better support collaboration. I asked their opinions on specific UI suggestions such as extra screens and a ‘what you see is what I see’ feature. Focus-groups were semi-structured to allow exploration of themes arising. I was especially interested to ask participants about collaboration, reference to objects, and their use of the virtual camera. I asked whether the camera made them feel they were ‘in two places at once’, and whether it made reference harder.

After the interview I paid external participants and they were free to leave.

The timeline of an individual trial session is illustrated in figure 15.

![Figure 15: Data collection timeline for individual trials in study 3](image)

6.2.6 Data Collection

I observed ten groups for a total of 22 participants. I recorded the participants’ screen video to allow for later analysis. Two participants’ recordings were excluded due to technical faults - I observed their sessions but had no video. Interaction analysis was performed on a cohort of
four experienced *Second Life* builders, ten with experience of other VWs or modelling tools, and six who were novices to 3d interfaces.

While observing lab trials I took hand-written notes, and occasionally asked participants about their activity.

Separately from the trials, I spent time in SL to discuss with experienced builders the themes that were emerging in the study. I posted topics to prompt discussion on the SL forum. Building classes and competitions were often held in the SL virtual world, and I participated in these to observe how experts were communicating about location.

Though most data were collected in a laboratory and some quantitative analysis made, the nature of study 3 overall remained qualitative and exploratory. My intention was to discover problems and solutions as SL users saw them.

Sample data from study 3 can be inspected in appendix F.

### 6.2.7 Data Analysis

I discussed my observations with other researchers in order to elucidate themes. I refined the themes that emerged in early trials by discussing them with the participants in later trials, and with users discovered in-world.

My qualitative data consisted of notes written during and after lab sessions, and transcripts from conversations held with expert builders in *Second Life*.

I combined each group’s individual screen recordings into split-screen videos which presented the coordinated viewpoints of the whole group, along with their voices.

A still from one of these videos is shown below in figure 16. It shows the simultaneous vistas of two participants performing the House task. In this picture, user 1 is ‘in-avatar’ and is watching user 2 move a gable into position. User 2 has moved their camera far behind and slightly above their avatar, though they can still see their avatar. Because user 2 is currently editing a prim (i.e., moving one of the gables), this prim is highlighted and the editing interface is visible on user 2’s screen.
The split-screen videos allowed me to review participants' interaction in detail, to discover problems that were not obvious during direct observation. I paid particular attention to problems of interpreting deixis. I transcribed and analysed fragments of conversation in the style of previous virtual worlds and CSCW work such as Hindmarsh et al. (2000) and Brown and Bell (2004), as described in section 3.2.2. This approach uses a simplified typography, with utterances listed in time order down the page, with a minimum of markup but with an emphasis on describing what users and their avatars were doing while the talk is taking place, and highlighting problems as they arose in the interaction. Selected fragments of trial transcripts that illustrate problems with interaction are shown in the Results section below.

In addition to interaction analysis, I made two simple quantitative analyses of the trials. First, I measured the proportion of time participants spent in each of the two camera modes. Second, I counted the deictic references used by each group. These measures were not intended to prove hypotheses, and it should be re-emphasized that the lab sessions did not have a formal experimental design. As discussed above, some conditions such as length of session were allowed to vary as participants saw fit. The quantitative measures simply allowed me to check patterns I had noticed during observation of the sessions.
I was also interested in the participants’ use of Second Life’s virtual camera feature, as both the Hindmarsh and Moore studies indicated that this was likely to impede users’ ability to deduce a collaborator’s viewpoint.

To make these quantitative analyses I marked up the trial recordings. I set up an Excel data sheet for each group, with time running down the page. An example sheet can be seen in appendix F. There were six columns for recording data. The first three corresponded to the group members. In these I noted whenever a participant switched between camera modes. The next three columns corresponded to the three deixis types: ‘with respect to avatar’, ‘with respect to object, and ‘by description of object’ (adapted from Levinson, 1996: see section 6.3 below). Whenever an event of interest occurred I added a new row to the sheet, wrote a timestamp on the left and recorded the event in the appropriate column.

The results were presented at a conference (Wadley and Ducheneaut, 2009) and then re-analyzed for this thesis.

6.3 Findings

I found that SL users did indeed encounter communication problems of the kind discovered in earlier research. But groups usually worked around problems by experimenting with different referential practices until they were able to convey their intended message. Participants were creative in their discovery and use of referential techniques, some of which involved unusual use of avatars and objects.

Bearing in mind that these were relatively simply construction tasks, groups rarely became mired in problems of reference for too long, and all completed their tasks to a reasonable level of quality.

The following exchange illustrates the kinds of problem that arose and the variety of techniques attempted. This group of three is performing the House task, and deciding how to position four walls to form the base of their house. The walls look identical and are currently positioned randomly in the vicinity.12

12 Each conversation fragment in section 6.3 is from a different trial. All participant names are anonymized.
A. So are you putting the walls together?

B. I’m moving one wall ... a third wall, towards the other two ... The one that’s tilting. [B marks a wall by changing its orientation in a fashion visible to her teammates.]

C. Oh, that was you! [A and C now know which wall B is editing.]

A. Why don’t you turn that over, and I’ll move the other wall? ['That' refers to the tilting wall, and ‘other’ to the fourth wall not discussed yet.]

C. Are you moving the one on the lower leftmost of the walls?

A. Well .. your left? [A and B laugh. C has attempted a spatial reference without realizing that A and B cannot dereference it.]

A. I’m going to move the one that I’m standing right next to. [A does not attempt to correct C’s attempt at deixis, but instead moves her avatar beside a wall in order to mark it to the others.]

A. The one that’s I guess kind of closest to [C] ... [A uses the position of C’s avatar as a reference point] ... why don’t we leave that one still, and then we can put the other three around it?

C. Mine’s above the ground. [C refers to the wall closest to his avatar as ‘his’ wall.]

A. That’s fine I think. Why don’t we just leave that one and put the other three around it?

In this situation the group tried several different referential methods. When one didn’t work they tried another. They made verbal reference to objects using words like “this” and “the other two” and to locations with expressions like “the lower leftmost”.

The fragment illustrates two creative methods of reference which were also used by other groups. One was to mark the objects themselves: in the above example B turns a wall on its axis and says “the one that’s tilting”. The other was to move an avatar so that it was standing beside the object being referred to. In the above example A moves her avatar to a wall and refers to it as “the one that I’m standing right next to”. These approaches were reliable, if unwieldy.

This method of marking-with-avatar was used by many groups during the Garden task, which required a ‘helper’ to frequently give directions to ‘worker(s)’. In the following fragment of dialogue, a helper is showing a worker where to place a table.

H: You see this one I’m standing next to? Don’t move this one – this one stays in place. [H is using his avatar to mark an object.]

W: Yep. So why don’t you just move where you want the others.

[Now H uses his avatar to mark locations.]
H: Yeah. [H walks to a different spot] The other one is going to go here - in front of me.

W: Right, hang on … [W moves the table]

H: And the last one is on the other side. [H walks around the house] Just about here.

An exchange from another group illustrates several reference techniques, including marking places with avatars, pointing with the avatar’s arm, pointing with the edit bar, marking an object by moving it, and verbally describing an object. At the point where this dialogue fragment begins they are confused about which object is the flagpole.

X: Is there a way to point? What’s the thing you thought was the flagpole?

Y: Hang on, let me just walk into it. See this thing that’s right near my hand? [Y is editing the house, so his arm is in the air.]

X: Which hand?

Y: Right in front of me. Can I point at it? [X places the pole in edit mode.] There we go. Why don’t I move it? If you’re watching it, I’m moving it back and forwards now. Can you see an object that keeps moving left and right?

X: Yes that’s the flagpole isn’t it? [… gap]

X: You just walked past a cement block. Are we supposed to do something with that?

Y: I think that’s the chimney.

Note that of the methods used in this example, pointing was the least useful. As in the example before, marking an object by standing one’s avatar near it, or moving it around, was the most reliable form of reference.

In general, participants displayed a broad range of building and communication styles, illustrating the variation that general-purpose systems need to support. Experience made a clear difference to participants’ ability to collaborate. Yet all groups were able to complete their tasks, albeit at different speeds and with varying quality.

### 6.3.1 Use of Second Life’s virtual camera

The ability to rapidly gain multiple perspectives of an object by moving one’s view-camera independently of one’s avatar location and orientation is not available in all virtual worlds. There is no doubt that this feature is useful to individual Second Life users while they are building. For example, when lining up the walls and roof of a house one must view the
progress from several angles. Having to walk one’s avatar around a house takes time, whereas detaching one’s camera to quickly look from different directions is more efficient.

However I suspected that this feature might cause problems for collaborating groups. Recall that Hindmarsh and colleagues found that users, although they could see each other’s avatars, frequently couldn’t deduce what another user was seeing, which caused verbal reference to break down. In the physical world, knowing where a collaborator is facing allows one to deduce what they can see, which in turn allows one to understand deictic expressions like “that thing there”. Hindmarsh et al. (1998) found that people failed to deduce from a collaborator’s avatar what the collaborator could see on screen.

As Moore et al. (2007) argued, a virtual camera which can be detached from the owner’s avatar should make this impossible, because it destroys the relationship between where an avatar is facing and what the avatar’s owner is actually looking at. The referential problems this can cause are illustrated by the following exchange. This pair performed their tasks but were plagued by a misunderstanding over viewpoints because one (here called ‘A’) stayed mostly in-avatar while the other (‘C’) detached their camera from their avatar and moved it to a vantage point above the house, where he looked down on the scene and seemed to ignore the avatars.

In this fragment the pair have assembled four walls and are about to place gables above two of them. Their avatars are at opposite ends of the house. However C’s camera is near A’s avatar, so that unbeknownst to A they are both viewing the scene from the same angle.

C: Let’s place those triangle things. [the two gables]

A: Where are those? Oh, the triangle things are around the front aren’t they? [It is not yet clear which end of the house is the front.] I’ll place the one on my side if you place the one on the other side. [This reference sets up the problem, as “my side” is an avatar-relative reference, but C is not looking from his avatar’s location.]

A: [after a pause …] I don’t know whether I’ve selected the same one as you. I’m selecting the one that’s further from the house. [The problem becomes apparent – they are both trying to move the same gable.]

C: Ok, do you see one moving? I selected one that I just raised up. [Again, moving an object was the reliable fall-back form of reference.]

A: Oh yes I see that one, ok good. I’ll pick a different one then. [pause] Oh you’re putting it on that edge?

C: I just put it on the nearest spot I could find. [The end of the house closest to A’s avatar is also the end closest to C’s camera, so they both chose the same spot to place their gable. There is noticeable frustration between the pair now.]
A: Where’s your character? [he means ‘avatar’] Oh ok, I see where your character is. I tell you what, can you put the gable on the house section closest to you, and I’ll move the one that’s closest to me? Unless you want to finish placing the one that you had. [A is still using avatar-relative reference while C assumes he means relative to viewpoint.]

C: Does is matter? I’m manoeuvring the one that I had.

A: Ok, I can move the other one I think. I’ll just walk around so I can see it better. [He walks his avatar to the other end of the house, where C’s avatar, though not his camera, are placed. During this move he apologizes for bumping into C’s avatar, though C seems to have forgotten he has an avatar.]

During the following exchange (Garden task) the same pair still struggle to understand each other’s references until user A asks user C – at this stage still ‘in-camera’ - to mark a location with his avatar.

C: If you’re facing the front of the house, you’ll see one in the front of the house with two chairs, to the left of the house is four chairs, and behind the house is two chairs.

A: So we’re going to treat me as facing the house right now? Do you want to see where I am?

C: Um, I see where you’re facing.

A: I tell you what, can you walk your avatar to what you’re calling the front of the house?

[C moves his camera back to his avatar, and walks to one end of the house.]

A: Ok. So you’re currently at the front of the house?

C: Yes I’m facing the front. [A proceeds to place the furniture in the correct spot.]

Being able to detach one’s viewpoint from one’s avatar, without other users knowing, clearly caused some referential problems. I was interested in to what extent users detached their cameras while working, and whether this affected either their experience of embodiment in the virtual worlds. For example, did SL users feel as though they were embodied in two locations at once?

Some participants, who were experienced MMOG users but inexperienced with Second Life, rarely detached camera from avatar, suggesting that extensive gaming experience may make disembodied viewing feel unnatural. Experienced SL users however said that maintaining both an avatar and a camera location did not bother them. On being questioned about “being in two places at once”, most said that this had never occurred to them. When asked, “How would you describe your location right now?” experts usually chose their avatar rather than camera location. One participant felt that the detached camera was simply a tool, and that while using it he continued to equate his avatar with himself. On the other hand, some felt that while they were building, their avatar was irrelevant and even got in the way.
One expert said that as a beginner she had identified with her avatar, but that over time she had begun to experience SL more as a building tool than a virtual reality. But her equally experienced building partner felt that to him, avatar location mattered, because avatars are how people find each other in the virtual world, and because: “something can happen to your avatar. You can get pushed or shot. Nothing can happen to the camera: it’s just a view of a picture.”

As described in section 6.2.7 (Data Analysis), I used the screen recordings to analyse how much time each participant spent in-avatar and in-camera. While analysing the videos, as participants moved between camera modes I recorded the time, then later added up how much time was spent in each mode. Because groups did not all work for the same period of time, I factored trial length out by converting the measure to a percentage of trial time. Averaged over all trials I found that participants spent about half of task time in each of the two camera modes.

While observing the trials I formed the hypothesis that time spent with camera detached from avatar correlated with SL experience. To test this I classified participants into three experience levels and calculated the mean time spent in-avatar for each cohort. Cohort A (n=4) were experienced SL users, while cohort B (n=10) were competent users, and cohort C (n=6) were novices.

The 6 means (3 expertise levels x 2 tasks) are illustrated in figure 17.

![Figure 17: Time spent with camera in in-avatar mode](image-url)
Comparison of means suggested that expertise, but not task, was a factor in camera-use. An ANOVA (using the Excel Analysis Toolpak) confirmed that Expertise had a significant effect on camera use (F(2,33)=8.93, p<.001). Expert SL users were more inclined to detach their camera, as one might expect considering this was an ‘advanced’ feature which was not available in most VWs.

Task did not affect camera use, and since House and Garden means differed in a non-linear manner between the three expertise levels I concluded that there was no interaction effect of expertise and task. These results surprised me somewhat, as I expected that the more intensive teamwork required by the Garden (helper-worker) task might have required different camera strategies. However this may represent two effects cancelling each other out. In the Garden task, helpers had to give directions to workers, and they frequently used avatar-relative directions, either with respect to their own avatar (“where I am standing now”) or to their worker’s avatar (“walk five metres forward, turn left and walk another five meters”). This may have caused both helpers and workers to spend more time in-avatar in order to decode the references. Counter-balancing this may have been the fact that the Garden task was performed after the House task, so that novices had more familiarity with the UI.

6.3.2 Verbal reference to objects and places

Natural language offers several ways to refer to objects and places. While analyzing the trial videos I counted references to objects and locations and categorized them according to the frame of reference used.

I used the scheme of Levinson (1996), who recognizes three frames which he calls relative, intrinsic, and absolute. In Levinson’s scheme, a ‘relative’ reference involves deixis from the speaker’s or receiver’s body’s point of view: I interpreted this here as avatar-relative reference. Some objects have their own ‘intrinsic’ frames of reference: for example houses may have an obvious front and rear to which other locations can be compared. Finally locations can be relative to an ‘absolute’ frame of reference such as compass points or a prominent object in the distance. I added a fourth category, ‘reference by name or property’ to count references which simply described an object, such as “the brown rectangle”.

Other researchers have used slightly different schemes: for example Herring et al. (2003) categorized references as “deictic”, “fixed unique” and “fixed non-unique”. Their ‘deictic’
category corresponds to my ‘relative’ category, while their ‘fixed unique’ corresponds to my ‘name or property’ category.

Although both Cartesian (x y z) and cardinal (north south east west) frames are available in Second Life, these ‘absolute’ frames of reference were rarely used by trial participants. On only two occasions, participants made use of Cartesian coordinates in order to describe locations to a collaborator. Only one group ever used cardinal directions. On one occasion a participant described a location as “the side closest to the sea”, which made use of a distant landmark and can be considered an absolute reference. Because it was so infrequently used, I excluded the ‘absolute’ category from my analysis.

Figure 18 illustrates the relative frequency of these forms of reference. ANOVA indicated that these proportions were consistent across groups (F(2, 27) = 6.37, p<0.01).

Neither task nor expertise level were significant factors, though expertise affected the overall number of references, with experienced participants making more references. While this appears to contradict a finding of Kraut et al. (2002), it suggests that those who were better able to handle SL’s particular style of representation and camera movement were more comfortable communicating about location.

The ‘relative’ frame of reference is of most interest. I suspected that knowing that collaborators were using the virtual camera might stop users from attempting avatar-relative references; however this was not the case: avatar-relative reference was used consistently across groups and tasks.
Note though that I did not attempt to record how many acts of deixis were successfully interpreted, as this was not always clear. Observation suggested while there were frequent problems, as described above, avatar-relative reference did often work.

It is noteworthy that even experts used deixis relative to a collaborator’s avatar despite knowing the collaborator might be in-camera and unable to interpret the reference. It seemed that experienced users were able to interpret with respect to their avatar’s frame of reference, even while their camera was elsewhere. They interpreted a reference such as “to your left” in the only reasonable way, meaning ‘to the left of your avatar’. If their avatar was out of view, they moved their camera to bring it into view, or returned their camera to their avatar, or asked for more explanation. Expert informants confirmed that this was only a problem in so far when they had to return to their avatar, they lost their camera position and would have to find it again. No participant ever checked with a team-mate, before using deixis, whether they were in-camera.

I asked participants whether they would like an extra screen permanently displaying their collaborator’s vista, but none of them thought that this would be useful. Some novice users suggested a ‘what you see is what I see’ feature that could transmit their vista to a colleague when asked for.

### 6.3.3 Non-verbal forms of reference

Having a humanoid avatar means that, in principle, a user can employ non-verbal forms of communication such as gesture.

*Second Life* offers a form of extended pointing whereby when a user selects an object for editing, their avatar’s arm points in the direction of the object, and a ‘selection beam’ appears between the pointing arm and the object. However reference using this technique was rarely observed in this study - though one expert claimed that it was used “all the time” in *Second Life*. Some participants remarked that although they could see a colleague’s avatar point while they were editing, the highlight on the object was only visible to the editing user. Avatar gestures other than pointing were never used in these trials.

On the other hand, moving one’s avatar to stand beside an object was a frequently-used form of reference. Participants also jiggled objects back and forth, tilted or moved them for collaborators to see, as illustrated above and as in the following exchange:
A: Are you rotating one of your walls?

B: Yes, is it rotating on your screen?

A: I’m rotating another wall. Yep, I just saw yours rotate.

At one point an expert participant changed the colour of a wall to mark it: however she said this was not a common technique because it is hard to restore the original texture.

Some experts said that it was common practice in SL to create a prim to mark a location, which was deleted after use: this is analogous to the use of marker pegs in physical construction. However no-one did this in these trials. It is possible that the technique did not occur to novices, and that experts did not find these tasks sufficiently difficult to require it.

6.4 Discussion

Like the study before it, study 3 examined user-to-user communication involving the defining feature of a virtual world: simulated three-dimensional space. While study 2 was concerned with the ‘macro’ spatiality of the distribution of avatars in a virtual world, study 3 examined the ‘micro’ level of collaboration around objects. At the macro level, a principle concern is knowing where one’s collaborators are. At the micro level the concern is knowing to which objects and locations one’s collaborators are referring when they speak.

The aim of this study was to answer the research question:

RQ3: How do groups communicate using voice while collaborating around objects at the ‘micro’ level of spatiality in a virtual world?

Spatial reference during object-focussed collaboration-at-a-distance was a known problem in early CVEs, and has been a significant topic of research (e.g. Hindmarsh et al., 1998; Kraut et al., 2002). I observed collaboration in a modern virtual world, allowing me to determine whether problems and practices had evolved since these technologies became mainstream.

I chose Second Life as the study system because it allowed users to manipulate objects and because it was a popular virtual world with a well-established user culture. I conducted a semi-naturalistic exploratory study rather than a formal experiment, but exploited the lab setting in order to observe users. I also employed ethnographic methods among the Second Life user base.
I observed problems with verbal references to objects and locations. However users devised interesting non-verbal methods of supporting verbal references.

### 6.4.1 Comparison with prior research

Some of these results can be directly compared to previous work. Hindmarsh et al. (1998) experimented with extended pointing, peripheral vision and a plan view. Second Life provides extended pointing via the 'selection beam', but users were ambivalent about the usefulness of this feature. Some users said it would be more useful if the object highlight seen by an editor was also visible to others, so that knowing who is working on what was more transparent. SL does not provide peripheral vision, and most participants felt that extra screens would be a burden. SL offers a plan view (the map) but none of these participants used it, probably because it displays insufficient detail to be useful when working at the level of objects.

Hindmarsh reported that stylized gestures were not useful for collaboration around objects, and study 3 supported this, with no use of SL's pre-programmed gestures observed.

These participants stated that an extra window showing their collaborator’s camera view would be superfluous. This accords with the finding by Fussell et al. (2003) that collaborators preferred seeing the shared workspace to looking through the head-mounted camera of a collaborator (see also Hindmarsh et al., 2001: p134-5).

Herring et al. (2003) found that novice users of ActiveWorlds ceased attempting spatial deixis when they found their colleagues could not dereference it. My participants were more successful with deixis, which may be due to better graphical representation in SL, or the availability of voice, which is better for quickly resolving ambiguity (Löber et al., 2006). It may also reflect people’s greater familiarity, due to the passage of time and increased use of videogames, with the representation style of virtual worlds. In some cases it was because users could exploit the ability to use their avatars as place-markers while moving their viewpoint elsewhere.

### 6.4.2 The 'out-of-avatar experience'

SL’s virtual camera trades off the benefit of rapid acquisition of multiple viewpoints against the drawback of making deictic references harder to resolve and foci of activity invisible. SL users effectively have two locations - their (public) avatar position and their (private) camera
position. This can foster deception (Irani et al. 2008), and some participants described a reduced sense of embodiment in the avatar, though they didn’t consider this a problem.

Expert users seemed to maintain a sense of where their avatar was facing, and used this to interpret references by others relative to their avatar. When this was impossible they simply returned their camera to their avatar to translate the reference, which can be done with one key-stroke.

It is possible, using scripting, to provide the location and gaze direction of one’s camera to other users. This can enable ‘you see what I see’. This feature is not available in SL’s standard UI, but was implemented by one entrepreneur as a software add-on which is available for purchase. I obtained this but found it had limited utility. One of the participants, an experienced SL builder, was familiar with this add-on but did not use it.

One might expect a detachable camera to diminish the relevance of the avatar (cf. Moore et al., 2007). Yet Yee et al. (2007) found that SL avatars obey physical-world proxemic norms of inter-personal distance and eye-gaze. For proxemic conventions to work, users must perceive each other as having a definite location and orientation in the virtual space. But in this study, participants often ‘parked’ their avatar while building, moving it to a socially appropriate position only when interacting with other users. In one session, a participant stayed in-camera until a new user appeared nearby, whereupon he went in-avatar and walked over to them. Others placed their avatars in a conversational orientation even when they had detached their camera. It seems that SL users regard their avatar as a public mediator of social interaction which can be ignored when not needed.

Arguably it is a form of ‘perception management’ to maintain proxemic norms using one's public embodiment while one's private focus is elsewhere. A user who maintains a conversational orientation while paying visual attention elsewhere must be aware that other users may be doing the same thing.

However to problematize detachment these users displayed from their avatars may reflect a bias towards believing that embodiment in an avatar is essential to the use of virtual environments. Perhaps what these users are telling us is that embodiment is not essential to object-focused collaboration – perhaps not essential to social interaction at all – and that avatars are not required in order to have a sense that other people are present in the environment. I expand upon this theme in section 8.3.8.
6.4.3 Articulating collaboration

*Second Life* can be viewed as a CSCW system in which users collaboratively create content. The common field of work (Schmidt, 2002) then is virtual objects and the space in which they reside. SL provides no specific mechanism for articulating work beyond standard text and voice tools, and the ability to view a shared virtual scene.

I was surprised to find that closely-coupled collaboration, in which two users work on the same primitive object at the same time, was rarely performed. On the contrary, the first impulse of many groups was to modularize their task. This may be because the articulation work required for close collaboration in a 3d environment represented too high a load. It might be easier to gain multiple viewpoints by moving one’s camera than receiving verbal feedback from a collaborator.

I implemented a worker-helper task in order to encourage more communication about objects and location. It is noteworthy that in other research where participants worked closely around individual objects, close collaboration was also 'forced'. For example, Pinho et al. (2002) required one user to move an object which was distant from their avatar, while another user closer to the object guided its placement. Roberts et al. (2006) implemented gravity so that two users were required to lift objects while a third joined them together. By contrast SL allows users to rapidly acquire a variety of viewpoints and does not implement gravity by default, so that objects can be lifted by a single user and will stay in place while the user works on other objects.

It may be that VW users will only collaborate closely around objects if physical-world constraints, such as gravity, and strict embodiment of camera within avatar, are reintroduced. But in a way the point of a virtual environment is that, unless simulation is required for training or other purposes, strict adherence to physical law is not required. Therefore it may be unrealistic to expect close collaboration of the kind we see around physical objects.

6.4.4 Representing a user’s focus of attention

Significant progress has been made in making avatar-relative reference more realistic, usually by improving the fidelity of pointing (e.g., Wong and Gutwin, 2010) or eye-gaze (e.g., Murray et al., 2009). However these extra degrees of freedom must somehow be controlled by the user. Alternative approaches suggested by this study include supporting verbal reference by
providing a publicly-visible marker which can be moved to mark locations, or by having the visual appearance of an object change when selected. Given the opportunity, this is what users choose to do. In other words, the VW should directly represent the user’s focus of attention rather than indirectly represent it via the gestures of their avatar.

In team combat games such as were examined in studies one and two, the avatars of fellow users are key ‘objects’ with which one is ‘working’. But for collaboration around objects, many users value being able to detach their view-cameras from their avatars, and will use their avatars as mere place-markers. If users are provided with objects to mark locations, embodiment in avatars may be superfluous.

6.4.5 Reflection on my methods

I employed a hybrid approach to study this research question: this may have been a strength and a weakness. I wanted the naturalism of ethnographic methods to exploit the fact that Second Life, unlike the older experimental CVEs, has a large user base with significant collective expertise. However the phenomenon being examined was a subtle one that required direct observation. Observation in-world would not have worked, because I could not see users’ screens, and because chancing upon collaboration-in-action would be rare. This implied the need to observe use in a lab.

To solve this dilemma I developed themes during lab observation, and discussed these with users both in-world and in discussion forums.

The lab setting itself can support both qualitative and quantitative methods and again I wanted to use both. But to compare participants’ performances quantitatively requires that conditions be tightly controlled and confounding variables minimized. In SL it is difficult to define the end-state of a building task, because prims don’t snap into place, and it is a matter of judgement when a group’s construction matches the experimenter’s intent. This makes it hard to measure completion times. I found in practice that users varied so widely in their style and approach that a purely quantitative study would have been impractical.

Allowing for a third member in groups required extra equipment and didn’t add many insights. In future work I would choose to observe only pairs.
6.5 Conclusion and next steps

I found that problems of reference to objects and locations discovered in early CVE research persisted in a modern mass-market virtual world. But users worked around them creatively. They did this not by using the avatar to point, but by detaching camera from avatar and using the latter as a place-marker.

While building, users spent about half their time detached from their avatar. They returned to their avatar to move it to a location they wished to mark, or to assume a conversational orientation when approached by another avatar. This suggests that, for object-focused work at least, embodiment in an avatar is not essential to the user’s subjective experience of using a virtual world. At best the avatar is a public face that the user constructs for social interaction: sometimes it is merely ‘parked’. This ‘virtual camera’ feature did not introduce as many coordination problems as one might have expected. This suggests that users were not trying to reconstruct their collaborator’s viewpoint.

These results challenge two popular notions discussed in chapters 1 and 2:

- that virtual worlds should resemble reality as closely as possible, and
- that avatars are essential to collaboration in virtual worlds.

My final study, reported in the next chapter, examined more generally the influence of voice on the user experience of Second Life.
Chapter 7

Study 4: Voice in a social virtual world

7.1 Background

My first and second studies demonstrated that voice communication could enhance sociability and real-time coordination of teams in game worlds. Studies two and three examined how voice could be integrated with the spatial nature of a VW, and the problems that arose when users tried to refer to points in space.

But study 1 showed that the extra social presence conveyed by voice introduced problems such as diminished ability to role-play, reluctance to communicate with strangers, and increased vulnerability to abuse. If the acceptability of voice in massively-multiplayer online games relies on a delicate balance between the benefits of team coordination and the problems of controlling social presence, this prompts the question: How might the introduction of voice influence the user experience of virtual worlds that are not games?

Second Life, the virtual world that was the subject of the study 3, is a social (open-ended) virtual world. It is not a game, and team competition is mostly absent. Instead Second Life’s users engage in a diverse set of activities that includes content creation, avatar customization, role-play, education, commerce and socializing. There is an emphasis on pseudonymity in SL, and exposure of one’s offline identity is proscribed. Many users feel themselves to be engaging truly in a ‘second life’ while using this system.

Linden Lab, the vendor of Second Life, announced in 2007 that it would introduce a voice channel later that year. This sparked a storm of controversy among SL users (Boellstorff, 2008; pp. 113-4). Many felt that voice would break the pseudonymity they valued, and ruin the atmosphere of this space that was so significant to them. Some threatened to quit Second Life if anyone used voice.

It seemed that voice was going to have an important influence on this popular virtual world. I felt this would be an ideal case for my final study. I could observe the influence of voice
communication from its announcement through its implementation and beyond. Because SL was not an MMOG it would provide a valuable comparison with studies one and two.

Second Life is discussed in detail in chapter 2. This VW appealed initially to role-players and then increasingly attracted interest as an inexpensive communication platform for tasks as diverse as business meetings, e-commerce and distance education. These are use scenarios whose communication requirements do not necessarily coincide.

For the first few years of its existence Second Life supported only text communication. There were two text channels: vicinity and instant message (IM). Vicinity chat was broadcast to any user whose avatar was in the local scene, and was not stored. By contrast an IM was directed to a specific person who needn’t be in the local area, or even logged on. (IMs received while not logged in were stored and presented on next login.) A screen-shot of Second Life, showing the UI for local chat, can be seen in figure 1 (chapter 1).

When voice was introduced to SL, two analogous channels were set up. The default spatial channel broadcast to users whose avatars were nearby – this was analogous to vicinity text chat. It was also possible to make a one-to-one or one-to-group “phone call” directed at specific users – this was analogous to IM, except that voice was not recorded for later access.

In 2009, two years after the introduction of voice, a Linden Lab press release\(^{13}\) stated that on the one hand, Second Life had become a major player in international telephony, with a significant proportion of the world's long-distance voice conversations made using SL’s voice-over-IP facility. Yet on the other hand only 50% of SL users had adopted the voice channel.

My fourth study was aimed at shedding light on the influence voice had on the user experience of voice in Second Life. Why did so many SL users reject voice and so many others embrace it enthusiastically? Why had there been such an intense controversy around its introduction? How did its influence vary according to the different types of use being made of Second Life?

\(^{13}\) accessed at http://lindenlab.com/press/releases/19_05_09 on 2nd December, 2011
7.2 Method

7.2.1 Aim

The aim of this study was to answer the research question:

RQ4: How did the introduction of voice influence the experience of different kinds of users in Second Life?

7.2.2 Participants

There were several participant groups corresponding to different data-gathering methods.

I recruited seven Second Life users to participate in interviews about their experiences with voice. Five lived in the Melbourne area while two lived elsewhere in Australia. Three were male and four were female. All were aged between 20 and 40.

These interviewees used Second Life for a range of purposes including online teaching, socializing, conducting business and creating art. I recruited them via advertisements in the Second Life user forum, two popular email lists ‘Second Life Research’ and ‘Second Life Education’, and the web-based Metaverse Journal. I focussed on participants local to me as I wanted my interviews to be face-to-face where possible; however I allowed participants to choose to interview ‘in world’ if they preferred.

I held a focus group discussion in Second Life on the topic of voice. I arranged this by approaching the convenor of a popular weekly discussion group which was well-known in the SL user community. They agreed I could host the discussion, and advertised it through their normal channels. Following Second Life convention these participants were pseudonymous; I knew their SL names but not their offline identities.

I spoke with and observed many SL users who I encountered during a period of two years.

I captured discussions of voice in blogs, mailing lists and the SL use forum. I did not record the identities of these informants.
7.2.3 Materials

No materials were provided to participants. I used an audio recorder to record face-to-face interviews.

7.2.4 Approach

I employed a variety of data-gathering methods during a two year period from 2007 through 2009.

I began by using an ethnographic, participant-observer approach. I joined Second Life early in 2007 and became a regular user in order to better understand the culture in this virtual world. I joined user groups, participated in discussions, bought items from other users, worked to improve my avatar’s appearance and clothing, learned to build, and explored the virtual world. While using SL I recorded voice and text conversations and took screen shots of interesting events. I kept copies of articles and blog entries relating to voice. During my first few months of use I kept a diary.

I joined two high-traffic SL email lists and monitored these, along with the SL user forum, for discussion about voice.

In November 2007 I convened a focus-group discussion about voice within the ‘Thinkers’ group, a long-running weekly in-world discussion group who met in Second Life. The participants included prominent SL users and bloggers.

In January 2008 I began interviewing users about their experiences with voice. There were 7 interviews in all, conducted either in person or in SL according to participant preference.

I used Second Life for a number of workplace activities over a period of two years. Participation in these activities required that I deal with problems of voice and text communication in my own work practice. This intensive period of participant research allowed me to develop insights and cross-check the experiences reported by participants. For example:

- During the first half of 2008 I helped teach and assess an architecture course at an Australian university (not my own) in which groups of students used Second Life to build projects.

- I conducted a demonstration involving students’ use of Second Life at a Melbourne secondary school’s Open Day.
• In 2008 I travelled to California to carry out study 3. During my stay I was a frequent user of *Second Life* and took part in numerous activities in-world, including meetings of a geographically-dispersed research group and of the organizing committee for an academic conference.

• I used SL to deliver a lecture, using voice, from California to an undergraduate class in Melbourne.

• During 2008 I took part in three mixed-reality music performances, which simultaneously took place in *Second Life* and a physical venue. (Reviews of these can be seen at http://massively.joystiq.com/2008/06/06/mixed-reality-melbourne-seventeen-unsung-songs, accessed on 7th December 2011).

• During 2009 I organized a weekly in-world seminar series on *Second Life* research, which utilized both voice and text and ran for several months.

My timeline of data collection is illustrated in figure 19. (Ethnography and reading of forums was spread throughout the research period.)

![Figure 19: Data collection timeline for study 4](image)

### 7.2.5 Data Collection

During my first few months of *Second Life* use I kept a diary of my experiences, taking screenshots and recording text and voice conversations for later analysis. For example, the following is a note I wrote after my first experience of using voice with other SL users. The note is illustrated with a screenshot of the region in which the encounter occurred (figure 20). The complete entry is shown in appendix G.
I just got voice working for the first time – I mean working as in I found some people I could talk with. … There were about 20 avatars there in a semi-open landscape – this is more than I’m used to seeing. I pressed the push-to-talk key, tried to say something, and suddenly I could hear conversation. … Three speakers were having a fairly intimate conversation … I could occasionally hear a young child crying or calling out in the background of one of the people speaking.

It took a good 20 minutes or so but I eventually figured out which of the avatars on the screen were speaking. In the meantime it was quite an odd experience – a conversation in my headphones that was quite unrelated to what I could see on screen, which was about 10 or 20 avatars standing around having typed-text conversation(s). …

At one point I was looking at three avatars who I thought must be the talkers, and one of them mentioned one of the other’s avatar-names, and then I knew I had found them. I listened to their conversation for a few minutes – this felt slightly odd, I was eavesdropping in a way, but didn’t really feel like I had connected socially with them. One of them logged out, then another. Then a couple of other avatars approached and joined in conversation with the remaining talker. At this point I outed myself as a voicer by saying something. They replied and were fairly polite, but it seemed awkward, like they knew each other and not me. … Using voice is a much stronger “social connection”, i.e. you feel a much stronger sense that the other people are people and not cartoon characters.

![Figure 20: Encountering voice users in SL (my avatar in foreground)](image)

My interviews with users were open-ended and lasted one to two hours. Most were conducted in person. However when requested, and for interstate participants, I conducted the interview in SL, using voice or text according to the interviewee’s preference. I transcribed the spoken interviews.

The focus-group was conducted using SL’s text chat system, as this was mandatory for this group. I saved the transcript for analysis.
I saved relevant threads from the official SL user forum, two popular mailing lists, and several prominent blogs. Voice was a popular and controversial topic in these forums.

Sample data from study 4 can be inspected in appendix G.

### 7.2.6 Data Analysis

I transcribed the spoken interviews, and used NVivo to code all of the data into themes. This analysis was presented at a conference (Wadley et al., 2009). While writing my thesis I re-analyzed the data and refined my coding hierarchy.

The final coding scheme emerged from the data but was also informed by themes that had emerged in prior studies. The code hierarchy can be inspected in appendix G.

There were 15 themes, and I sought high level concepts within which to categorize them. One was the choice of voice or text, specialized uses for voice, how choices were made and the controversies it engendered. Other high-level themes were media richness and its effect on online identity as well as griefing. I grouped together themes concerning technical properties of voice, and the ability of voice to breach the boundary between online and offline communication contexts. My final coding hierarchy can be seen in appendix E, figure 42.

These results are presented in the following section. Each theme is illustrated with fragments of data, following the style suggested by Glaser and Strauss (1967). (See section 3.6 for discussion of presentation style.) I discuss the themes in greater depth in section 7.4.

### 7.3 Findings

This section begins with findings that cluster around the SL user’s ability to choose voice or text, and variation in the responses made by *Second Life* users to the introduction of voice. Not only was there significant controversy over voice, but particular uses made of SL that suited or did not suit voice.
7.3.1 Controversy over voice

User feelings about voice were far more ambivalent in *Second Life* than in the MMOGs I examined in study 1. I found that the introduction of voice elicited extreme reactions in forums and discussions. Opinions ranged from "It hugely enhances SL for a lot of people" to "I hate voice so badly. I just hate it. I want it to go away". The controversy confirmed the observations of Boellstorff (2008, pp. 113-4).

I hate voice. I hate it with a passion. I'm a woman and I'm shy. I am a nerdy bookish person and I'm more at home with text. It's a place where my nasal voice and softness disappear and my ability to write lets my personality really come out. This is a real loss for me.

On the other hand, some felt that extreme pro- or anti-voice positions were inappropriate and that users should be left to choose the medium they preferred:

Fine, if you don't want to use voice, just don't use it! But I know that for many applications voice is important. So I am happy to have the option available, and the freedom to choose using it or not.

Likewise one of my interviewees felt that voice worked in some situations, with text always available as a fallback, and was happy to let others choose the medium when conversing.

There was speculation about how many people were using voice in SL. Many felt that it was restricted to isolated pockets or to certain types of users. A forum poster reported: "I jumped randomly around the grid looking out for the famous white dot indicating voice usage. Among hundreds of avatars I saw not one who was using voice." Another said: "I don't run across people using it at all, other than club DJs."

The discussion participants expressed similar scepticism about the popularity of voice, such as: "Predominantly, everyone uses text", and "I use it now and again and some people I know use it now and again but the vast majority of people who I meet don't use voice most of the time."

A Linden Lab employee, during their in-world 'office hours', said that: "People do like voice. Some people really don't, but a lot of people really do."

One informant, echoing Fulk et al. (1990), felt that it was conversation-groups, rather than individuals, who chose modality: “Group decision. Sometimes governed by tech. If everyone has voice we use it. If one person doesn't we all go text.”

This controversy illustrates the variety of types of user and use found in this virtual world, and the power of the new communication medium to radically change the user experience.
7.3.2 Specialized use-cases suited to voice

Informants described certain uses of Second Life to which voice was particularly well-suited.

One of my interviewees attended regular meetings of an interest group that used SL voice, and another used SL with voice for business meetings. An informant who used SL to telecommute said that: "Voice is a lifesaver. It makes meetings much better." Several informants were involved in education, and all of these felt that voice was good for online teaching. One interviewee said:

   I use both voice and text. Has been fabulous having the option to use voice chat added to SL. Particularly for education. Many educators I work with didn't 'get hooked' until voice was added.

Another of my interviewees used SL for teaching a foreign language, and felt this was an ideal application for voice, as he could both type a word and speak it. In this scenario he found no problems with voice other than technical ones. A discussion participant commented: "There is a language school in SL. They *have* to use voice." A mailing-list poster agreed: "Like all language teachers, I imagine, I am delighted that there is voice in SL".

A popular activity among some SL users was transmitting live music into SL through the voice channel. A forum poster remarked: "At Help Island Public there was a guy giving a musical performance using voice, playing guitar and singing." A discussion participant noted: "A choir playing live in SL. That's hard to do in text!"

Apart from live music, several other forms of performance were enhanced or enabled by the addition of voice to Second Life. One of my discussion participants listed scenarios to which voice was essential: "performances (theatre, poetry and storytellers, comedy acts, classes to name just a few!)." Another commented: "I just came from a play rehearsal that couldn't have happened without voice." A forum poster discussed a group of actors who staged plays in SL.

   I don't use voice most of the time myself, but the plays that the SL Shakespeare Company performs at the Globe are in voice, and they are wonderful. The audience is reminded several times to turn voice ON and turn their microphones OFF so that they don't make noise during the performance.

Some forum posters felt that voice was well suited to content creation in SL: "I am with a group of builders working on a quite big project. I usually use voice for our meetings and other communications." A blog poster commented: "I like to use voice when I am hanging around and talking to friends while building, because it frees up my hands to continue creating things."

One blog poster commented on the use of voice for cybersex: “As for who is using voice, it's escorts!”
Clearly the range of purposes for Second Life use exceeded that of the MMOGs of my earlier studies, and for some of these voice was either an enabling or a disabling technology.

Sections 7.3.3 to 6 concern the properties of the voice modality and how these influence user experience.

### 7.3.3 The richness of voice

Participants reported that speech conveyed more richness and subtlety of meaning, more quickly than does typed text, and were often positive about this. One interviewee (who chose to be interviewed in text!) praised "the immediacy of the connection, and the intensity of the bonds/emotional commitments", and reported that “voice _does_ help with understanding some of the communication overtones”. A forum poster felt that: "Text is often too formal, some kinds of meaning can't be conveyed in text." Another felt that voice "makes it much easier to present complex concepts”.

Especially when explaining something to a newbie it is faster and easier to speak. Meetings with friends and group members are more efficient when most people are talking instead of typing.

However one interviewee felt that the ambiguity of text had subtle advantages, offering an interesting analogy from the voice vs text distinction to the movie vs book distinction.

I think that [voice] broadens, but reduces the depth of the communication. This is the difference between a movie and a book.

When you have a book, the reader has to synthesise all the images. This process acts to bond those images and impressions deeper into the user/reader's world-view. When the images are provided from an external source, the user is simply a consumer, not a co-creator. This to me is the real strength, power, and danger of a visual + text environment.

*Books and movies are quite different user-experiences. Does voice make that much of a difference in SL or other VWs?*

In a relationships/communication medium (such as SL is for a large number of participants), the addition of voice can be both powerful, and restricting.

In other words, communication by text required more use of one’s imagination, and for some social interactions, especially those involving role-play, this could be a good thing.
7.3.4 Identity and anonymity

In concordance with the findings of study 1 and prior research, informants felt that voice conveyed more information about the identity of the person speaking than did text – information such as their gender, ethnicity and age. This interfered with the anonymity which had previously characterized the virtual world experience and which many users valued (see e.g. Raybourn, 2002: p253), and which some exploited to enable the role-playing of alternative identities. Loss of anonymity was one of the main reasons cited by SL users for rejecting voice.

An interviewee commented on protection afforded by the relative social leanness of the text medium: “You don’t feel so exposed in text.”

Some felt that voice had robbed Second Life of its atmosphere.

SL’s original silent world is more magical without the crassness of voice and interruptions by real humans.

For me SL is all about creating the image that people want to be, and to be honest I really would rather imagine how some people talk than to actually hear them.

One role-playing interviewee, a member of the ‘furry’ sub-culture, complained that voice suited better the ‘augmentationist’ style of SL user who merely wanted the virtual world to be a communication medium, and less suited to the ‘immersionist’ who wanted to be in a fantasy universe.

Voice makes SL a better telephone to chat with their real-life friends. But what use is voice to a furry?

On the other hand, a discussion participant felt that voice was suited to immersion in the space because as a form of communication it was more immediate than typing. “Using voice, you could be ensconced in headphones, not needing a keyboard between you and what you want to say. That sounds MORE immersive to me.”

A forum poster pointed out that actors and offline role-players successfully pretend to be other people while using their natural voice. “I was in the Society for Creative Anachronism for years and was quite able to maintain immersion in the medieval world in ‘first life’.”

Gender-play – using an avatar of opposite gender to the user - seemed to be as popular in Second Life as it was in MMOGs. Confirming Yee (2005) and Blinka (2008), gender-play usually involved males with female avatars. However while I did not measure this quantitatively, I did form an impression that in Second Life the use of opposite-gender avatars
was more likely to be exploration of gender identity, than was the case in MMOGs (as reported by Blinka and MacCallum-Stewart and confirmed in study 1).

For example one male interviewee who liked to use a female avatar said he was apprehensive about the introduction of voice into Second Life, to the point of becoming anxious and depressed. This participant stated that he was not gay or transsexual and was married with children, and that his female SL character was very important to him. He described his use of a female avatar not as ‘playing a female character’ but as ‘being’ female while in SL.

Regardless of the reason that SL users chose opposite-gender avatars, they reported the same problem that MMOG players did: that their offline gender was readily exposed by their speech.

I also have many SL friends who role-play, and don’t want to reveal identity with their voice. I think it depends on what you’re role-playing. For example I have several friends who are male but choose female avs. They obviously don’t use voice.

Clearly this had the potential to cause major problems for some SL users. One informant related an instance of voice disrupting gender-play and precipitating the user’s voluntary exit from the social group.

One female at the meeting was outed by her attempt to use voice as being male in real life, for no one’s benefit. She had been quite believable before that as a female, and I haven’t seen her again since that meeting. As far as I have been able to tell, she left SL shortly after that.

Conversely, for female users, voice could cause problems by giving away the fact that they were female. “What turned me against it was the fact as soon as guys would hear a girl’s voice it was like a feeding frenzy.”

On the other hand, one female forum poster said of voice: “I use it to prove I’m a girl. I find that in some situations I get treated better if I voice rather than not.”

One male interviewee who created content in SL using a female avatar, and who communicated by text, felt he was treated better by a particular client who thought he was a young woman.

It was widely known among SL users that projection of offline identity was a popular reason for disliking voice. As a result, some users who refused voice (for whatever reason) encountered suspicion from other users who assumed they must be trying to conceal their identity. “You get the oh it’s really a man, woman, kid. People assume all the wrong things when someone doesn’t use it not when they do.”

One text-user reported being expelled from a piece of land by its voice-using owner who assumed the visitor must be there under an assumed identity. Another suggested: “I think there
needs to be an accepted community standard way of saying ‘I don’t use voice’ which is accepted as a choice and not thought of as suspicious.”

Some users preferred transparency of identity. An interviewee who used SL for conversing with business partners and customers said she liked voice as it gave her a better idea of who she was dealing with. A forum poster reported attempting to join a group who “insisted I come online and chat with them by voice: it would help them decide whether I was trustworthy.”

Some preferred to use voice only when they already knew the people with whom they were speaking. One interviewee “would tend to approach new people with text first in social settings”. A forum poster reported that: “I keep voice on all the time, but typically only use voice with family and friends.”

One interviewee complained that national accents could make utterances impossible to understand: a problem unique to voice.

There was some discussion of different types of users in SL. One informant referred to: ”two different SL audiences - recreational and business, with very different uses and needs. Business prefer / need voice.” Several commented on what was felt to be a widespread belief that ‘immersionists’ prefer text and ‘augmentationists’ prefer voice. And some forum posters saw a connection between Linden Lab’s introduction of voice and the company’s management of these two user types, believing that Linden Lab wanted to attract business and education users at the expense of their early adopters.

I suspect there’s always been tension between the people who want to extend their RL selves into the virtual space, making it like ‘Skype-3D’, and those who want to take something other than RL there. […] In the early stages of Second Life it was the second group that were hooked and were consequently paying; but in order to finance the next stage of its growth Linden Lab need to provide features which interest the first group at the risk of upsetting the second.

### 7.3.5 Personality and modality preference

Several informants felt that a user’s personality contributed to their preference for voice or text. “I find that a lot of people who don't like voice are shy and lack confidence. I wouldn’t say this is the case for everyone, but I think it's a pretty common reason.”

One interviewee felt that everyone was anxious about the sound of their own voice. “People are insecure about their voices - they don't like the recorded sound of their own voice, worse than seeing their own photo.”

Some felt that voice allowed offline extroverts to dominate the virtual world.
If you introduce sound the whole ambience changes: the shy are revealed as shy, and the noisy start
to dominate. (It is hard to type LOUDER than everyone else in the room!)

Quiet people, in my view, tend not to enjoy having to listen to a bunch of loudmouths talking over
each other, or having to figure out a way to get what they say into the discussion. Easier just to clam
up and resign to being a passenger, tune it out and go back to your Sudoku or blog-browsing in the
other window.

An interviewee who used SL for teaching said: "Text is sometimes much more appealing for
those who are a bit more introvert, whilst voice can enable the performers amongst us." She
felt that "sometimes text is a more effective mode of communication and participants will
share more."

However one informant felt that, as their confidence had increased, they had come to embrace
voice.

Over the past months I changed my opinion completely and went from a strong anti-voice standpoint
to a very positive one. IMHO voicing is a matter of self-confidence. When it was introduced I was
afraid of people making fun of my accent or that they won't understand me, or that guys would
constantly harass me. Both things are just not true. It's all about confidence! I started using voice
with close friends only but soon moved on to using it whenever I can.

7.3.6 Multi-tasking and multiple conversations

While concepts like ‘immersion’ tend to convey the idea of a user being completely absorbed
in their current activity in a virtual world, in fact it seems that many SL users multi-task: either
combining SL use with offline activities, or by engaging simultaneously in multiple
conversations within SL.

Informants felt that text was better suited to multitasking and multi-conversing. An interviewee
said: "especially when you're multitasking as I often am, text is easier." A discussion
participant noted that:

When you're on a phone call, all your attention is focused on the call. That's normal: voice demands
high bandwidth. Whereas on text, you can multitask pretty easily.

An educator praised the ability to multitask during class time: “I WANT my students to be
texting, IMing and generally doing bunches of other stuff.”

A forum poster appreciated the ability to hold multiple conversations without interlocutors
being aware of the fact that they were multi-tasking.

When your main mode of conversation is text, you can talk to lots of people at once, about different
things, and the participants don't all know you're talking to other people. When it's voice you all talk
about the same thing, in public, usually with the same people.
By contrast, an interviewee reported being confused about who was talking when he heard several people using voice at once.

It can be difficult to have a meeting or any interaction with a lot of people present as one person's voice tends to cut through someone else's and people don't naturally take turns when talking.

Voice in games suffers from the inability of users to see each other and use visual cues to decide who will speak next.

A poster to the SL forum preferred text over voice for its ability to support multiple threads within large conversation groups.

I prefer IM and email over phone calls and voice conferences with people I'm doing business with, in person, any time there are more than a couple of people involved in the conversation. It's got nothing to do with my identity or whether I'm role-playing or not. It's simply that voice is not well suited for many-to-many conversations. It tends to break up into multiple 1-1 conversations (where the cocktail party effect comes in) or it turns into a series of 1-many presentations.

Some participants discussed the simultaneous use of voice and text. It is common in SL discussion groups, conference talks and similar forums for a main speaker to use voice while other attendees type their comments and questions. “I have lead discussions where I have deliberately used both modes […] Perhaps presenters use voice to share at the beginning of session then ask people to work in groups with text.”

A forum poster reported that:

in both SL and in academic Skype conferences I have recently found myself in private side-chats, discussing how and why the main thread was going wrong, planning an intervention with some allies, and then intervening.

On the other hand, one interviewee found voice to be confusing when speakers were also IM-ing each other. Some forum posters agreed that it was hard to use both modalities at once: "I personally don’t talk much because I stop doing everything else”, and "My problem is that I can't do computer stuff and talk at the same time".

Clearly multi-tasking of various kinds is common among SL users, and the text medium supports far better than does voice.

In a similar vein, informants reported that the larger the group, the greater advantage text had over voice. This accords with studies one and two, and with the findings of Lober et al. (2007). One interviewee said the largest group in which he had used voice had six users, and that he’d felt like he was talking in a crowded room. A discussion participant said:

Voice works well in small intimate groups and is very good at one-to-one's. Its efficacy diminishes as the group gets bigger.
The finding that many SL users multi-task is an important challenge to the idea of immersion in virtual worlds, which I discuss in greater depth in chapter 8.

Sections 7.3.7 and 8 concern properties of the text medium and how these influence the user experience of Second Life users.

**7.3.7 The processability of text**

Informants valued the processability of text (Dennis et al., 2008) and made use of this property when conversing in this medium. A text conversation could be saved to disk and read or searched later. Second Life prefixed each utterance in a text conversation with the name of the user who typed it, making it easy to interpret later. SL also made it easy to capture a conversation via a menu option. Analogous features were not available for voice.

I greatly prefer text chat for a meeting. Everyone has full transcripts of exactly what was said and by whom, and what was agreed to. No misunderstandings. In voice, there's no way to record or transcribe the conversation, or to identify who is speaking.

Another interviewee agreed: "It's easy to keep minutes of a meeting if it's in text". A discussion participant felt that "text will persist just because it is easy to remember what was said". Another noted: “I attended a lecture where the speaker used voice. And, in that context, it was really good to HEAR her rather than read her. Shame I have no transcript though.”

These experiences emphasize the range of uses to which Second Life is put. Taking minutes and re-reading lecture transcripts are obviously not scenarios likely to be encountered in MMOGs.

But even in regular group conversation, text storability of text presented advantages. As in studies one and two, informants commented on a text channel’s ability to scroll back over past conversation. An interviewee said: "With text, I can go back to the log and follow various threads of chat".

Whereas voice channels required careful turn-taking, text channels did not, as individual utterances were simply placed in a queue and could be easily discerned. A forum poster reported: “I find that typed meetings go faster, because you don't have to wait for people to finish talking, you don't have to repeat what you just said, you don't need 'rules of order'.”

Some informants reported making use of the opportunity to edit a text message in the short gap between composing and sending: “Sometimes I go to type something but then delete it, because I realise I should not say it.”
This echoes findings from studies 1 and 2. Whereas text can be misinterpreted more easily than voice, it gives users the opportunity to check and edit potentially provocative statements before they are sent. The immediacy of the voice medium by contrast encourages users to broadcast without thinking.

### 7.3.8 Avoiding the need to type

Most people speak fluently, but few are competent typists. This fact provides a clear advantage to the voice medium. “I use voice whenever possible. I usually don't mind whether or not the other person types or talks, as long as I can avoid having to type :)”

An interviewee said that "typing skills makes a huge diff", but felt that younger users who had grown up with computers might cope better with text: "I have a young nephew in the family. It is fascinating watching him grow up in the same way. He can type his name at three." A blog poster commented: “I find that the worse someone's typing skills are, the more inclined they are to use voice.”

Another commented on the overloading of manual controls, a major problem already discovered in MMOGs. “My favorite use is for keeping hands free while exploring the world with a friend, since it's the only way to walk and talk simultaneously.”

Some commented on accessibility problems associated with typing. “One of my family members in game has cerebral palsy which makes typing difficult for her.”

(Of course, there are other disabilities which prevent use of a voice channel. However these did not surface in this study.)

These findings regarding benefits and disadvantages of the text medium show that it was not merely the properties of voice that drove users’ choice of medium.

Sections 7.3.9 and 10 concern mundane technical problems that plague voice users, a recurring theme in my work which is discussed in greater depth in chapter 8.

### 7.3.9 Technical problems with the voice channel

The theme of users struggling with technical problems with the voice channel featured strongly in studies one and two, and arose again in this study. Problems of this kind are mundane, and
arise from technical constraints rather than the nature of voice itself; however they are clearly of importance to voice users.

At its most basic, a problem can arise if a user simply doesn’t have the extra equipment required to send and receive voice. People do not always use their own computer for VW use, and may to access SL on non-voice-equipped PCs. Carrying a headset so that one is ready for Second Life use on any PC one might use during one’s travels does not appeal to all users. “Odds of me having a headset and mic hooked up to use voice? Not often.”

Problems with reliability, equipment setup and poor sound were reported by a large number of SL users: “The technicalities of voice in SL are complex if you haven't used voice apps before. [...] For me it's mostly the tech issues that stop me from using voice”. An email-list poster said: "The quality leaves so much to wish for. Generally speaking other people hear me well, but I usually hear others rather soft and often scratchy and distorted.”

One interviewee - a lecturer who had taught in SL using voice - reported that "every tech failure that could occur did!” Another interviewee spoke of difficulties convincing IT departments to unblock the required IP ports at institutional firewalls: "Voice is a toughie for lots in edu. Different proxies, and the hardware - nightmare".

Voice users are dependent also on other users’ having the right gear and knowing how to use it well. Again, as in studies 1 and 2, the problems of getting volumes of a group of voice users to match was mentioned: "A smooth voice discussion also requires everyone to have their voice settings at similar levels”. One forum poster frequently encountered “a mic turned up so high that the sound is redlined”. A poster to an email list commented:

For it to be a pleasurable experience for everyone, you all need to be using headsets and microphones and have your own levels adjusted appropriately. As soon as one person decides to use speakers or whatever, the whole experience can go to pot. Similarly, someone who has the input level on their microphone set way too high (or, sometimes, boosted as Windows allows) is going to ruin it for everyone. This is not an issue with Second Life voice, but rather an issue with anything involving live user-generated audio.

Users holding discussions in SL reported requiring a setup period during which participants debugged each others’ voice hardware:

The levels-setting tends to happen during the meeting, which can be a pain because you need to get the meeting done in the time period allotted - you can't be mucking around with technical matters.

A forum poster reported the following experience while holding a meeting in SL:

One user was consistently too loud or overmodulated, and therefore hard to hear. One user was so soft voiced that she was almost always impossible to hear, unless you were standing right by her shoulder. Two others were unintelligible, no matter what they did. Five minutes into the meeting, I reverted to typed chat, and abandoned using voice for the rest of the meeting. I had to cover all the original talk all over again, because a lot of critical points in the first part were lost or garbled.
The larger the group, the more likely there was to be one user who was unable to use voice. “If there are ten avatars in a room and nine of them have voice, I will use text to make sure that tenth person isn't excluded.”

An educator noted that students in particular, who often were irregular users of SL, often did not have the right hardware or know how to use it in order to use voice. “There are accessibility issues there in that often I have a mixed [voice and text] crowd. I always get a transcriber to summarize what's happening.”

It was easy for a user to unwittingly ruin a group conversation by accidentally setting up a feedback loop between their speakers and microphone. Two informants spoke of this problem:

People won't always know they're causing feedback or echo for other users. Politely inform them how to fix this, by either isolating the microphone from their speakers or by using a headset.

NO ONE should go around using voice through their SPEAKERS. All everyone else hears is echo. Of course the offender doesn't hear this part, so they think all is well. Please, please, invest in a headset with nice ear hugging earpieces.

Furthermore the Second Life client software requires significant resources, and voice adds to the burden that a user’s computer must handle. “Some PCs simply can't handle it. I myself wasn't able to run it because of my sound card set up. […] Now that I can start voice it will crash after about 20 minutes.”

One interviewee felt that the hardware form factor for voice use was a problem: “Perhaps the problem is the hardware design of the headsets – these should change so they look less like a telemarketer's headset, perhaps more like the minimal mobile phone headsets.”

Clearly the problems of having access to the necessary hardware and getting it to work properly are significant for many SL users, especially those who are casual users or who need to access SL from a variety of computers.

### 7.3.10 Problems with spatial audio

Most SL users operated voice in its spatial or proximity mode. Like the system examined in study 2, this arrangement simulated the transmission of sound in air by attenuating volume according to the distance between avatars. Spatial voice is designed to allow users to structure their conversation groups in a natural way, yet this process was not seamless.

One interviewee felt that while people have a good understanding of how far their voice travels in the physical world, the rules of spatial voice transmission in SL are less intuitive. (This
Another interviewee was concerned that they couldn’t control transmission distance by changing the volume with which they spoke.

Voice is not realistic enough - can’t really shout or whisper like in RL as the distance/loudness/volume relationship is too lofi. The fall-off is determined regardless of whether on top of mountain, inside a little room, voice is not altered by SL surroundings.

Another complained that: “Voices don't sound like they are coming out of the avatar's mouth.” Possible this is because while voices of distant avatars are made quieter, they are not frequency modulated to sound like distant voices (in which bass and treble frequencies are attenuated).

Furthermore, by mimicking the physics of physical-word voice transmission, spatial audio can reintroduce physical-world problems to the virtual world. One forum poster noted that if they attended a public lecture or large meeting, it was hard to be in a location where they could hear all the speakers at once. They might position themselves to hear the main speaker clearly, but would then be unable to hear an audience member at the back ask a question. SL had no “megaphone” to project voice over a large area when required.

Another feature of voice is that the transmitting medium – sound – is difficult to contain. This allows spoken utterances to be overheard by eavesdroppers for whom they were not intended. Sections 7.3.11 through 13 concern the problem of controlling who hears sounds.

### 7.3.11 Physical world sound leaks into virtual world

As noted by a range of participants in studies 1 and 2, an unintended side-effect of using microphones is that sounds other than the user’s voice could be unintentionally transmitted into the virtual world.

I don't do voice, and keep it turned off. […] The last couple of people I listened to on voice (I was text chatting), not only were hacking their lungs out at me, but making some other disgusting sounds as well.

Another said: "Most the time it’s some mom screaming at the kids with the mic on, someone having their music on very loud in their RL house”. Another complained about "sex talk, sorry don’t need to hear others moaning and groaning”.

Some informants reported maintaining awareness of which sounds in their environment they might unintentionally transmit, and adapting their practice to suit.

I don't do voice. Too high an ambient noise level in the room, with my kid and/or my mate watching TV or playing music in the same room. Not to mention a barking dog and other distracting noises. I set myself up so I could hear their voice chat on my computer's speakers, and I typed my replies.
Some were concerned that transmission of environmental sounds would cause privacy breaches. “I have family at home and I'm constantly being called by name to do something. That's one major thing I don't want broadcast over the internet.”

An interviewee who created audiovisual art in SL was concerned that voice would recreate a problem that plagued galleries and music venues in the physical world: audience members chatting and disturbing the experience of fellow attendees.

### 7.3.12 Virtual world sound leaks into physical world

Similarly, some were wary that sounds arising in or intended for the virtual world might annoy people co-present in their physical surroundings: "I usually do not use voice, mainly because it is distracting to others in the house." A forum poster complained: "I have the problem of other house members being disturbed by my loud booming sounds." Another noted: “Hell, maybe you have a baby sleeping next to you. Or a housemate who doesn't want to listen to you talk into a mic while she's watching a movie.” A parent remarked:

> I'm not much of a TV fan, but I've found an acceptable "alone together" compromise where they watch and I play, and we occasionally interact. Headphones and microphone don't really mix in that environment.

This problem was also reported by MMOG users in study 1 and can be thought of as a general concern for VW users.

### 7.3.13 Eavesdropping

The leakage of sound between virtual and physical worlds raised the problem of eavesdropping. Two kinds of eavesdropping could occur. In one, people co-located in the user’s physical context could eavesdrop on the user’s SL conversations. This overlaps with the concerns in section 7.3.12.

> They may have family members in close proximity or be in circumstance where talking to a computer isn’t private enough.

On the other hand one interviewee was not concerned about her SL conversations being heard in her physical surrounds.

> Are you ever using SL in (physical) situations where using voice would be heard by strangers around you?
Yeah, because I work in SL, I regularly use it on the road. And am used to people looking at me funny while I talk to my computer. For worky stuff it doesn't bother me.

But there was also the possibility of eavesdropping within the virtual world, by listeners whose avatars were near the speaker's avatar but of whom the speaker was unaware.

My partner and I regularly talk to each other in-world using voice. And although I have an eye on the mini-map and an awareness that people can listen in I still do it.

This problem is unique to voice systems that use spatial propagation. A ‘radio’ system broadcasts to a channel with a specific membership. Spatial systems however broadcast to whomever has an avatar close enough to the speaker’s. The broadcast radius is not visualized (in Second Life) and so it is not always clear which avatars in the vicinity are transmitting one’s utterances to their owners.

This problem is made worse in Second Life because of the virtual (detachable) view-camera, which was discussed in some detail in study 3 and by Irani et al. (2008). A user can position their camera independently of their avatar location, and its position is invisible to other users. Furthermore the user can choose to listen from their view-camera, which makes eavesdropping within the virtual space easy and undetectable.

People can zoom in from a long way away to listen in. For newbies all of that is very hard to understand. I run some activities inworld with participants to experiment with how far away for chat and voice – it is interesting to see reactions.

The ‘leakage’ of sound between and within offline and online contexts had a major influence on the usefulness of voice and the user experience of voice users. This result is particularly significant taken in conjunction with my findings on SL users’ desire to multi-task and to engage in multiple conversations, activities that therefore become much more difficult when voice is used.

Finally for this study, sections 7.3.14 and 15 return to the problem of misuse of the channel, a problem created by the properties of voice discussed above such as enhanced projection of social presence.

7.3.14 Griefing

Some people try to make the Internet experience unpleasant for others, and a voice channel creates new opportunities for doing this. One interviewee claimed to have switched SL voice on when it was implemented, left it on for ten minutes, then turned it off after hearing only verbal abuse. She felt that griefers were worse in voice than in text. Another claimed that voice
is primarily used for griefing. “The only places I really see it in use are at the Welcome Areas - and usually to harass and belittle others or say racist/sexist stuff.”

A discussion participant who role-played in the 'furry' sub-culture was concerned that: "Every time I hear someone do voice in SL I just hear people insult me and say Furfag and the like".

Informants offered different explanations for why griefing was worse when spoken rather than typed. Some felt that an insult "becomes more personal when voice is used". Some thought voice attracted people who were disposed to griefing: "Says something about people using voice I guess". Some felt voice exaggerated offline communication problems:

The girls stop talking completely, the shy people shut up mostly, and all that is left are the 12-18 year old guys, and it becomes a locker room.

A blog poster reported that the voice channel was used for a variety of offensive noises other than voice:

Go to a welcome center sometime to see what voice is really used for in SL. Although there is sometimes useful conversation most of what I hear are mics keyed to proudly express various bodily noises and mics placed next to speakers emitting annoying and offensive music while the avatar basically looks like it is afk, ie it is intentional.

As a participant in study 2 pointed out, it is possible that griefers prefer voice because it is harder to be caught. Whereas text conversations can be easily logged, recording the voice channel is harder, and a recording of sound is not so easily traceable to the offender. It is unlikely that Linden Lab systematically monitors the SL voice channel.

Voice lent itself to a particular form of griefing reported by one informant:

The wildest game I’ve seen played with voice chat is when someone records ‘choice phrases’ from someone, uploads the WAVs, and plays them back while the person is still there. […] Don't voice drunk, sarcasm is your enemy, and watch what you say around welcome area crowds.

Despite these reports, one interviewee at least had a more positive experience: “I haven’t had any problems with griefers or interruptions when using voice.”

Clearly problems with abuse of the voice channel will depend on individual users and the people they happen by chance to encounter; however it appears that it is a significant problem for many.

7.3.15 Need for etiquette

While griefing represents deliberate misuse, informants recognized that some misuse was accidental. Some felt that, over time, the development of etiquette would improve voice
practice. An interviewee remarked: "I think we're all still working on the social niceties of what's ok and when". A forum poster said “There should be classes on how-tos and voice etiquette”. A blog poster wrote:

The mishmash I see in current voice-use is, I believe, simply the growing-pains of learning how and when to use a new technology. This problem will likely solve itself as people come to realize the many shortcomings of using voice as they have been.

However these optimistic predictions did not seem to be shared by “anti-voice” Second Life users.

### 7.4 Discussion

‘Social’ virtual worlds such as Second Life are technologically similar to massively multiplayer games, but are culturally distinct. When voice became popular in MMOGs it was perhaps inevitable that it would soon be introduced into social VWs. This did indeed occur, but voice provoked controversy and has been embraced only by a subset of users. This study has helped to elucidate why this is the case.

Social worlds are used by a wide variety of people for a wide variety of purposes. The diversity is greater than in game worlds, because in a game user actions are constrained by rules and the imperative to compete.

In study 1 I identified that the most compelling use-case for voice in game worlds was coordination of team members who are dispersing quickly through virtual space. This scenario is largely absent from social worlds. Yet there are features which are common to both types of VW: for example users are represented by pseudonymous avatars, they negotiate physical and virtual contexts, many are engaged in identity/gender-play and so on.

The aim of study 4 was to answer the research question:

RQ4: How did the introduction of voice influence the experience of different kinds of users in a ‘social’ virtual world?

I chose Second Life as the test system because it was the most popular of the available social worlds, featured a rich user culture, and had been adopted for instrumental use such as teaching, commerce and meetings. I used multiple methods to gather data about the influence of voice on the SL user experience. These included interviews, group discussion, online ethnography and participant observation. My data were the reported experiences of SL users,
triangulated against my own experiences as a regular user for two years. This acquisition of data from multiple sources conformed to the principles of grounded theory and UX research.

### 7.4.1 How this work built on previous studies

Study 4 tied up themes which had developed throughout my program of research.

In study 1, I found that voice strongly influenced the user experience of massively multiplayer games. Key phenomena included sociability, team coordination, channel clutter, the broadcasting of social presence and offline identity into the virtual world, and the leakage of sound between online and offline contexts.

Study 2 examined a different kind of voice channel, again in a team game context. Some of the themes of study 1 just mentioned re-emerged here. I discovered also that spatially-propagated voice promoted immersion, provided cues about the location of team-mates, acted as a relevance-filter for crowded channels, and afforded new gameplay strategies that tied to movement in space. Eavesdropping in the virtual space became possible –this was interpreted in the context of the game as espionage, which users found to be an interesting tactic that enhanced gameplay and immersion.

Study 3 examined the communication that takes place when people collaboratively work around objects in a virtual world. At this spatial scale the location of one’s collaborators is usually clear, but the objects to which they are referring often are not. Users employed standard forms of deictic reference, including avatar-relative references. But instead of using avatars to point or gesture, they used them to mark places.

To round off my project it was necessary to examine the influence of voice on virtual world use that did not involve a team combat game. Study 4 therefore repeated the approach of study 1, but within a social virtual world. *Second Life* was the most popular of these systems, and was already attracting researcher attention, allowing for comparison with the results of others. It offered ‘spatial’ and ‘radio’ communication in both voice and text modalities, and boasted a diverse population of users engaged in a range of activities. In *Second Life* many of the pros and cons of voice discovered in study 1 arose again. However the lack of a need to coordinate moving teams, and the greater diversity of users and use-cases, surfaced a number of new advantages and disadvantages of voice.
7.4.2 What study 4 unearthed

Study 4 highlighted the diversity of uses being made of Second Life, as well as the wide range of attitudes held about how SL ought to be used. These illustrate the importance that SL holds for many users. The study confirmed that voice is differently suited to different situations, purposes and personalities. It is clear that communication modality matters to Second Life users. Yet the criteria by which they judge modalities are many and varied, and cannot be reduced to a single concern.

It is important when seeking to understand communication in this system to recognize that users operate within a complex superimposition of social contexts, in both physical and virtual space. Everyone has a physical context, in which they are differently audible and visible to others. When a person logs into Second Life they enter also a virtual context, which is just as complex. Other users can move in and out of communication range simply by positioning their avatars in virtual space. Exactly which other users are in listening distance is not made clear, and one can move one’s “ears” and “eyes” far from one’s avatar without this act being known to others.

Some of the problems and preferences of SL users discovered in this study represent a rediscovery of the results of prior research into communication media. Other concerns are uniquely important to virtual worlds and would be unpredictable from research conducted outside the VW context. For example, pseudonymity and role-play are of importance to many users. There is evidence for the existence of distinct types of users (or types of use) which have different communication requirements.

The preponderance of technical problems encountered by voice users in VWs has rarely been addressed in prior research (an exception is Nilsson et al., 2002, who studied voice-enabled meetings in ActiveWorlds). However problems with sound quality and setup are a major concern for users and most certainly do influence choice of modality on occasion.

7.4.3 Properties of media

Many of the opinions, both positive and negative, which SL users hold about their communication media are related to the seemingly higher information-richness of voice relative to text. Users in both workplace and recreational contexts commented favourably on the ability of voice to efficiently convey subtlety of meaning and to reduce misunderstanding.
But voice also conveys information about the person speaking, and even other people in their physical context. While some SL users value this enhanced social presence, it is also the source of most of the controversy about voice.

To a greater extent than is the case with most online systems, users of social worlds like Second Life are likely to directly encounter and converse with people they do not know and will never know offline. Exposure of personal characteristics through voice is exploited by some as a way of getting to know colleagues and customers, and forming bonds of trust. On the other hand, people seeking to use virtual worlds pseudonymously, especially those whose avatars do not match their personal characteristics, object to voice precisely because it transmits such information. Furthermore, people who reject the richer medium may be held in suspicion because others assume they are trying to conceal their identity. The information richness of voice also intensifies the psychological impact of grieving.

Second Life users recognize that personality factors such as shyness can play a role in people's willingness to use voice, especially when conversing with strangers: this confirms the findings of study 1 and of Stritzke et al. (2004). A related issue is that some individuals dominate the voice channel, which is compatible with the prior finding that voice transmits status cues more than does text, and magnifies offline hierarchies (France et al., 2001).

The results of this study confirm that the framework of media synchronicity theory (Dennis et al., 2008), which analyses media along multiple independent dimensions, is relevant in the virtual world context. My augmented model for the MMOG context, presented in study 1 as figure 9, is supported by findings from Second Life. Users are aware of the differential performance of text and voice with regard to supporting transmission velocity, symbol sets, parallelism, rehearsability and reprocessability, and structure their use of media accordingly.

SL users engaged in text chat accept that there will be a delay between each round of a conversation. They use this to more carefully edit messages, which provides an opportunity for impression management. They also use the delay to multitask, for example by taking a phone call, doing computer work in another window, or engaging in other conversations. Text users appreciate the ability to undertake multiple simultaneous conversations: this cannot be achieved so easily in voice. Users appreciate the ability to glance back through the log of a text conversation while discussion is in progress, and later save the log as a transcript, and they have discovered useful ways of using text and voice simultaneously in large discussion groups, where frequently the norm is for one person to speak while the others type. The practice of text-messaging in a back-channel while a presenter is speaking is widely accepted in SL.
In practice, SL discussions are often held in text for mundane reasons. Voice is less reliable than text chat. In a large enough group, someone will probably lack voice hardware or have it incorrectly configured. If one user cannot use voice, most groups will revert to text, lest that user be excluded. The best solution is often a mix of several media (supporting the suggestion of Watson-Manheim and Belanger, 2007) and it is interesting to note that workplace-oriented systems such as Wonderland emphasize this.

### 7.4.4 Managing multiple contexts

Accidental transmission of environmental sounds into *Second Life*’s voice channel can damage privacy, create embarrassment, and break immersion for other users. This was found also in studies 1 and 2, where participants unexpectedly found themselves listening to their fellow users’ family members or work colleagues. As noted in prior research, eavesdropping and accidental transmission are problems for any voice medium, and researchers have observed active management of this problem by telephone users (e.g. Faulkner and Culwin, 2004).

Accidental transmission of sound is especially problematic for *Second Life* users, for a number of reasons. SL transmits sound based on proximity, and it is not obvious to users which avatars are in hearing range. The *Second Life* UI makes intentional eavesdropping easy, by allowing users to listen from their view-camera position rather than their avatar position. Eavesdropping within the virtual world is a problem unique to users of spatial voice in spatial systems. And while the word ‘eavesdropping’ implies the interception of a secret message, it is sometimes not so much the content of the message that the telecommunications user wishes not to divulge, but the fact that they are engaged in multiple conversations. The concern of the multi-tasking or multi-conversing user may not be so much one of information privacy as of saving face.

The physical context of SL use is often a user’s home or office, where friends, family or co-workers might overhear voice transmissions. Unlike telephone calls, which are typically brief, a SL session might last for hours. Many SL users avoid communicating by voice in the virtual world if it might disturb those in their physical context, or if the sound of conversation, music or television in their physical context might be broadcast into the virtual space. Despite the fact that users are often considered to be occupying an artificial space separate to the real world, the fact that they are situated in the physical surrounding in which they conduct their daily lives is of fundamental importance to their experience (cf. McCarthy and Wright, 2004; Taylor, 2006).
Some people use VWs to role-play fictional identities, and some prefer to use an avatar of the opposite gender. Preserving the boundary between offline and online identities can be a complex problem. Different friends and family members, and online and offline colleagues, will be differently aware of the identity play being conducted. Voice, by broadcasting information about offline identity, and merging online and offline conversation contexts, makes management of a pseudonymous identity a complex task.

Finally, some people perceive a social stigma attached to virtual world use, adding extra pressure not to be overheard while using these systems.

### 7.5 Conclusion

Study 4 demonstrated that Second Life users - like other communicators - rarely want to project maximum information about their identity and activities. Users are situated in a complex combination of physical and virtual communication contexts, and lean media help them cope with this. I conclude that media richness is a contestable design goal for virtual worlds, as it has proven to be for other forms of CMC.

Much of the appeal of recreational virtual environments stems from the ability they afford to play characters in a fictional world. Even in environments without a back-story such as Second Life, users frequently play characters that are idealizations of their real selves (Ducheneaut et al., 2009) and are concerned to control the amount and type of social presence they convey.

The immersionist-augmentationist divide can be understood as a spectrum of different preferences for media-richness. Even augmentationists do not wish to project *everything* about themselves and their actions into a virtual world. Immersionists want to be 'in' the virtual world, and in it with other people, but do not want to 'be themselves' in it; and do not want the virtual world to closely simulate the physical. They don’t want to be anonymous – they want a public identity that persists over time and with which other people knowingly interact. They put significant effort into maintaining this fictional identity over long periods of time. I propose the term “fictional social presence” to label this seemingly paradoxical desire to project the social presence of an identity other than one’s own.
8. Discussion

8.1 Overview

Virtual worlds have become successful as recreational spaces and there is renewed interest in using them for computer-supported collaborative work. Implementing voice communication in these formerly text-only systems has significantly influenced their users’ experience. Yet prior to the work reported in this thesis there was limited research into how communication modalities influence the user experience of virtual worlds. This led me to adopt the research question:

RQ: How does voice influence the user experience of virtual worlds?

Using methods from the research traditions of UX, ethnography, conversation analysis and grounded theory I conducted four studies that examined communication by voice in a variety of scenarios and configurations, in order to understand the benefits and drawbacks of voice, and the new forms of collaboration that it has enabled. I found that voice radically changes the user experience of virtual worlds.

In this final chapter I reflect on my project as a whole, how it was conducted and what it contributes to knowledge. Finally I discuss the limitations of my research methods and suggest opportunities for future research.

I begin by revisiting the contributions of individual studies.
8.2 Contributions of individual studies

8.2.1 Study 1: Voice in MMOGs

Study 1 confirmed that MMOGs are complex social environments, in which people conduct diverse activities to which communication by voice is differently suited.

Voice provides two major benefits to MMOG users: making the game world a more convivial place, and helping them achieve success in the game. It allows real-time coordination of tactics and precludes the necessity to use hands simultaneously to communicate and to control avatar movement. Voice can be preferable to text in situations requiring negotiation, such as team management and the distribution of loot.

The problems of voice include poor sound quality and channel outage. Users sometimes have difficulty knowing who is talking, do not know whether they are being heard themselves, and frequently need to spend time debugging their own and other users’ setups. Users are more sensitive about speaking than they are about typing, and more concerned with who can overhear them. Abuse is perceived as more impactful and more likely to happen. Shy users prefer text, are less communicative when voice is adopted by their team, and are more easily dominated by extroverted team-mates. Voice is more suitable to permanent than ad-hoc groups, and after adopting voice some group members become reluctant to play with strangers.

Voice projects more information about the user than does text, interfering with identity- and especially gender-play. Voice is unsuited to asynchronous communication, and is prone to congestion, forcing large groups to split across multiple channels or to permit only high-ranking members to speak.

These results are partly prefigured by prior research conducted outside the virtual worlds context, such as the interplay of media richness and online identity (Ma and Agarwalm 2007), the influence of personality on media choice (Stritzke et al., 2004), the desire to engage in multiple conversations (Jones, 2002), the effect of group size (Lober et al., 2007), vulnerability to eavesdropping (Faulkner and Culwin, 2004), the differing abilities of media to be saved, stored and searched (Dennis et al., 2008) and the question of whether users know each other (Carlson and Zmud, 1999). However some of the aspects of VW use that are influenced by voice, such as role-play and coordination in virtual space, are unique to this context.
8.2.2 Study 2: Spatial voice

Study 2 examined a voice system that simulated the propagation of sound in air, finding that while the spatial metaphor represented a restriction over ‘all to all’ propagation, users discovered subtle benefits and warmed to it over time. This was the first reported study of use of a spatial voice system.

Propagating voice spatially adds socially relevant information to messages: an example of social translucence (Erickson and Kellog, 2000). Users receive not only spoken messages but a sense of how close and in which direction is the speaker. In collaboration where the rapid negotiation of space is key to success, information about other users’ positions is a valuable resource.

Messages such as orders, warnings and requests for help often refer to actions and events that are local to the speaker. Thus a spatial channel acts as a filter for relevance and reduces channel clutter, as individual users can only receive messages from people who are nearby in virtual space. However because the volume of an utterance depends on many variables, it is not the reliable indicator of distance that users expect.

Using spatial voice, members of opposing teams can hear each other, and this enables new strategies. Users taunt opposition players, spy on them and give false information to fool enemy spies.

Despite these advantages, some users wondered whether the older ‘radio’ mode voice systems would be better for teams who were disinterested in immersion and simply wanted to win. This would mean that the choice between these metaphors represents a choice of competitive advantage over a realistic environment similar to that taken by World of Warcraft players who use software mods to replace their realistic view of the space with direct depictions of their team-mates’ state (Golub, 2010).

8.2.3 Study 3: Collaboration around objects in Second Life

Study 3 examined the verbal and non-verbal communication that takes place during collaborative building in a voice-enabled virtual world, comparing results with those from earlier research on experimental CVEs. I found that problems of mutual understanding remain in the modern systems, but that users are creative in inventing referential practices that allow them to work around these problems. For example users mark objects by moving or tilting
them or changing their colour. They create temporary place-marker objects, to be discarded
after use. They use their avatars to help convey reference to objects and locations, not by
pointing or gazing at them but by moving the avatar to the place in question and referring to
“where my avatar is standing”.

The fact that in Second Life, view-cameras can be detached from avatars, makes deduction of
a collaborator’s viewpoint unreliable, yet people still use avatar-relative deixis, indicating that
their collaborators can dereference these enough of the time. Many abandon their avatar while
engaged in object manipulation, returning to it only for social interaction. Users are not
bothered by ‘being in two places at once’.

These results suggest that object-focused collaboration might be better supported by making
objects themselves more accountable, and that embodiment in avatars is not essential to the
user experience of virtual worlds, but rather is situationally dependent. This contributes to
research aimed at supporting reference during collaboration in virtual worlds, as well as
research into the implications of embodiment in avatars and the relationship between users and
avatars.

8.2.4 Study 4: Voice in social worlds

In study 4 I examined the influence of voice on the user experience of Second Life. This virtual
world is quite different from the MMOGs examined in study 1. It is used for a much broader
range of activities, including instrumental purposes such as teaching, commerce and meetings.
Second Life is devoid of game activities such as team combat, which studies one and two and
earlier work showed to be a key scenario favouring voice.

I was able to confirm that different types of users have different requirements and operate in
different circumstances, so that one communication configuration may not suit all. I also
detected a clustering of attitudes that supports the categorization of SL users as ‘immersionists’
and ‘augmentationists’, or along a continuum of immersion.

Many SL users either love or hate voice. It can be a boon during meetings, discussions, online
teaching, collaborative building, performance of music and drama, and virtual sex. Some users
feel that speech enables richer and more nuanced interaction, and is superior to text when
discussing complex topics or helping fellow users. Business and education users like the
ability of voice to convey information about a speaker’s identity. However the easy immediacy
of voice promotes communication without contemplation. Voice is problematic for people who
engage in role-play, and this problem is critical for many SL users. Some of these dreaded the introduction of voice into Second Life and even contemplated quitting the virtual world.

SL users negotiate a complex superimposition of virtual and physical contexts, and like to multi-task and “multi-converse”. Using voice makes it harder to prevent the projection of evidence of one’s activities and surroundings into online communication contexts. This focus on context is in line with recent approaches in HCI.

Choice of communication medium is important to SL users. The criteria by which they judge media are many and varied and cannot be reduced to a simple, single concern. Some of the problems and preferences expressed here represent a rediscovery of the results of earlier communication research outside the VW context, while others are uniquely important to this context.

8.3 Contributions of the thesis as a whole

I now present conclusions which emerge from the project as a whole. These are derived by comparing the cases to each other and to the prior research reviewed in chapter 2.

8.3.1 Voice transforms the user experience of virtual worlds

Studies one and four showed that communicating by voice radically transforms the user experience of virtual worlds. As social presence theory suggests (Short et al., 1976), voice projects more information about the identity of users into the virtual world. This can make the experience significantly more sociable in some situations, but interferes with role-play and pseudonymity in others.

Issues arise from the complex technical requirements of voice-over-IP, the ability of users to multitask and conduct multiple conversations, and the necessity to manage conversation in both a physical and a virtual context. The existence of different sub-groups of people with different attitudes to use adds complexity to this picture: my research showed how these attitudes map to choice and use of media.

Virtual world use involves people in real settings projecting constructed identities into fictional spaces. User-to-user communication lies right at this intersection of fantasy and reality. The suitability of different communication media to a particular situation of use depends on a
dynamic compromise between the need for people to understand collaborators, and their desire for privacy and identity-play. While voice offers much, text will persist as the best option in some scenarios.

Prior research into mediated communication can shed some light on communication in virtual worlds, but research conducted within VWs is required for a fuller understanding. My work contributes to design by emphasizing user experience, and that simply adding voice to a virtual world and ‘hoping for the best’ will not do: users may reject the channel and even reject the virtual world.

8.3.2 ‘Virtual media’: simulated telecommunication in VWs

There is a virtual distance between any two users of a virtual world which is independent of the physical distance between them, but which must, like physical distance, be overcome by a communication medium in order for them to converse. While VWs are in one sense telecommunications media, they are unique in that they locate users within a 3d space and provide them with a “medium within a medium” through which to communicate linguistically. The VW system not only simulates a space but a means for delivering messages within the space. Therefore I consider the message-delivery mechanisms used in virtual worlds to be ‘virtual media’.

Using this term reminds us to differentiate the mechanisms provided for linguistic communication from the virtual space itself. That is, it would be a category error to compare virtual worlds to other communication media, in the sense that one might compare, say, World of Warcraft to email. WoW comprises both a virtual space and a collection of virtual media with which to communicate within the space.

The term ‘virtual media’ also highlights the fact that the mechanisms currently used within virtual worlds typically emulate existing media that are in use outside the VW context. Voice and text channels are usually set up to emulate either radio (e.g. Ventrilo or raid chat) or sound (e.g. spatial voice or vicinity chat). Virtual worlds typically have an ‘instant messaging’ feature that is like regular IM systems, and is somewhat like email in that if the recipient is logged out, the message is delivered upon next login. These are examples of virtual media.

However virtual media could implement arbitrary transmission mechanics. Metaphors other than sound and radio could be used. For example, voice channels could emulate mobile telephones: by this metaphor, communication would be available at most places in virtual
space (except where the user is out of range). Each communication session would have to be deliberately initiated, and the phone number of the receiving player known. Alternatively a voice channel could emulate land-line telephones, in which case communication would be available only at particular places, and might cost ‘money’. (Wadley et al., 2005).

So as well as examining how well a virtual medium supports communication, attention might also be paid to whether the physical medium it emulates suits the VW’s setting. For example the sound-in-air metaphor would suit a game with a medieval setting better than would the radio metaphor; perhaps better still would be a simulated landline telephone presented as a magic ‘speaking stone’. Access to virtual media could be a privilege that needs to be earned by achieving game goals: for example players might need to find a radio, just as in many games they need to find weapons.

### 8.3.3 Media richness and virtual worlds

In chapter 2 I reviewed critiques of media richness theory (MRT) which demonstrate that while all communication media are in a sense replacements for face-to-face communication, the usefulness of a medium cannot be reduced to the fidelity with which it emulates the face-to-face condition. Studies 1 and 4 demonstrated that the controversy over the introduction of voice to virtual worlds partly reflects disagreement over the desirability of projecting social information into them.

Schroeder (2010, p. 6) argued that the study of virtual environments provides clues about the nature of media richness. I agree, and would add that the post-MRT message has not always been heard in the VW context. This may be because some people believe that the purpose of a VW is to be a maximally rich medium. Indeed the history of virtual environments has been imbued with an imperative to make them ever richer, and to make a user’s representation more closely resemble their offline reality. For example Roussel and Gueddana (2007) suggested that the goal of immersive environments was “to make the communication more natural, more intuitive and more realistic”. And Saeed et al. (2008) discovered that educators’ intention to use Second Life was predicated on its offer of rich communication.

The imperative to offer fidelity to the face-to-face condition is also visible in the drive to develop input technologies that reflect VW users’ movements in their avatars. For example, Schroeder (2010) argued that the evolutionary end-point of workplace-oriented CVEs will be avatars that accurately reflect their users’ movements, gestures and facial expressions.
Arguably, this imperative is represented in the introduction of voice: note how many SL users perceived this as reality intruding upon the fantasy world they had constructed.

However my findings emphasize that the key outcome of post-MRT research – that richer doesn’t necessarily mean better, or in other words that face-to-face is not necessarily the gold standard to which communication should aspire – applies to virtual worlds as much it does to other communication contexts.

It is unlikely that VW users want a maximally-rich medium. Many of them rejected voice because it projected too much of their offline presence into the virtual world (study 4), and even enthusiastic voice users playing MMOGs recognize a range of problems related to their voice systems’ unwanted projection of social presence (studies one and two).

This view accords with Golub’s (2010) critique of the idea that people become immersed in virtual worlds because of the realism of the simulation. Golub based his argument on two pieces of evidence. One was that people had proven that they could be ‘immersed’ in text-based MUDs which have no visual fidelity at all. The other was that users of beautifully-rendered MMOGs are willing to use mods which replace the 3d scene with numbers and charts that explicitly represent the underlying game state, to gain an advantage in the game. Some players also turn the game’s immersive sound effects off so that they can concentrate on gameplay and on each other’s conversation. Golub argued that it was “this commitment to the group project of raiding, rather than sensorial immersion in virtual worlds” that brought about immersion in the MMOG. Analogous group activities in non-game virtual worlds would include building, commerce, in-world travel, and discussion groups.

Critiquing media richness approaches in the VW context has implications for the design and governance of these large systems. Technical advances which attempt to link users’ in-world representations more closely to their physical selves could be seen as attempts to increase the social presence that is projected into the space. However we can extrapolate from the controversy over voice to predict that many users will resist such devices, because they do not want to project maximal social information.

For example, although I did not examine reactions to gesture-tracking devices, I predict a response similar to that stimulated by the introduction of voice will be likely if these are introduced to mass-market virtual worlds. Some people will adopt them enthusiastically. But the knowledge that one’s every move and gesture was being transmitted into a virtual world would probably prevent some users from adopting gesture-tracking technology. Those who did adopt it would likely need a new kind of ‘mute button’ in order to prevent transmitting
accidental gestures, such as scratching, raising cups and resettling oneself in one’s chair, because these are not relevant to one’s actions in the virtual world, and could break privacy.

My research highlighted the fact that people like to multi-task while they are communicating with others who are not co-present, or when they are accessing information through the Internet. A lightweight, lean medium, which transmits only what the sender intends and can be used asynchronously, is more conducive to conducting multiple activities and conversations than is an immersive medium which locks the user into one task. Immersion is desirable in entertainment contexts (Badique et al., 2002) but may be a bug, rather than a feature, during instrumental use. Perhaps this underlies the relative commercial failure of virtual worlds compared to ‘2d’ technologies such as the web.

This conclusion contributes to communication theory by extending the critique of MRT to virtual worlds. I demonstrated the desire that VW users have to control the projection of their social presence. Many choose text over voice in order to prevent the transmission of audio cues to their actions and context. Many exploit attenuated social presence in order to role-play. There is a tension between people’s desire for presence and co-presence and their desire to control what is transmitted to others about their offline identity.

8.3.4 Fictional social presence

To extend the discussion of social presence, there is no doubt that people use virtual worlds in order to “be there together” with others (Schroeder, 2010). This was a strong result throughout my research and a consistent finding in prior work. Consider for example the prologue to Nardi (2010), where the author relates the excitement she felt during her first WoW session upon realizing there were real people in the world with her.

In that moment I became aware of other players. I was not alone in the Night Elves’ Garden of Eden but surrounded by real human players who would interact with me. I was touched that another player had given me something for free, without my asking or even having a way to thank him. (Nardi, 2010, p. 5)

However while people enjoy the presence of others, and want to project social presence to them, the identity they wish to project is not necessarily their own. Many VW users want to be in the world, but do not want to be themselves in it. They want an identity that persists over time and yet is constructed, and they put significant effort into maintaining this identity. I use the phrase fictional social presence to describe this persistent, yet fictionalized, identity that some users wish to project into virtual worlds.
This is the paradox of role-play. Users desire interaction with each other, but are constructing their identities in a fictional setting, and wish to protect their offline identity. Using voice instead of text to communicate makes this a more difficult enterprise.

My study participants felt that there were very few ‘true’ role-players, especially if that term is taken to mean someone who attempts to ‘speak in character’ while using voice. Some MMOG players only group with people they know offline, and project their regular offline persona to them (e.g. the participants in study 2, and one of the groups in study 1).

The two ideal types known as immersionism and augmentationism in the SL community may not explain much actual use. I conclude that most people adopt an approach somewhere between these extremes. To varying degrees, people exploit the social opaqueness of media as a resource in their interactions with others (Erickson and Kellog, 2000). They understand which portion of their bodily and situational reality is on display and which is not, and use this knowledge to construct a performance (in the sense of Goffman, 1959) in order to achieve social goals.

Performance construction is particularly visible when VW users engage simultaneously in multiple tasks or multiple contexts: the next two sections describe these phenomena in detail.

8.3.5 Virtual worlds afford multi-tasking

VW users are aware that their use of lean media affords others less opportunity to monitor their offline activities. They exploit this to multi-task and even ‘multi-context’, engaging in impression management and multiple conversations with a range of people both online and physically co-located (c.f. Jones, 2002).

This need not imply deception of a malicious kind. VW users may feel themselves to be present in a shared virtual environment and cooperating honestly with collaborators, yet at the same time maintain an awareness of what is and is not being transmitted into their virtual context about activity in their physical context. They use this ‘reality gap’ to engage in mundane activities such as answering the phone, talking to others who are co-present, using other software, or eating. Such activities need not be accountable to their fellow VW users for effective cooperation in the VW to take place.

Multi-tasking may illustrate a politics of attention (Turner and Reinsch, 2007) in which different agents in an interaction have different agendas. Teachers and team leaders, for example, typically demand more attention to the task at hand than their subordinates desire to
give. Lean media allow the latter to give the impression that the correct amount is being paid. Thus different degrees of media richness may be desired by different actors in the same interaction.

It is sometimes said that virtual environments should be as immersive as possible, that using a VW should be like being transported to another place, and that VWs succeed when users are so absorbed in the simulation that they forget their physical surroundings (Biocca et al., 2003). But most communicators need to multi-task. An employee using a VW at the office, a parent engaging simultaneously in family life and Second Life, or a teenager playing a game while chatting with co-present friends, cannot afford to be so immersed in the virtual world that they ignore the physical one.

**8.3.6 Virtual world users manage multiple contexts**

Studies 1 and 4 showed that a significant problem for a VW user is their need to simultaneously manage physical and virtual contexts of communication, and that this influences their choice of modality and vice-versa.

Research by Dourish (2004) and others has emphasized the importance of the physical and social contexts in which a technology is used. Users of communication technologies in particular engage in a complex relationship with their surroundings. While speaking or using a telephone, people are aware of who in their physical context might be able to overhear them, and they tailor their communication to suit (Madell and Muncer, 2007). Online and offline activities and identities intertwine as people multitask in their physical and virtual contexts (Aarsand, 2008). This focus on context aligns with the ‘third paradigm’ of HCI research:

> Where the second paradigm down-played whether an office had books in it or that a computer sitting under a desk produced lots of heat when analyzing mouse performance, that the third paradigm recognizes that “externalities” are often central figures in the understanding of interaction. (Harrison et al., 2007).

Virtual worlds take the problem of understanding context a step further, because VW users simultaneously occupy locations in physical and virtual space. A user will be seen or heard in virtual space according to the VW’s transmission rules (Benford referred to the zone of visibility/audibility as the user’s ‘aura’), but it is not always clear where the boundaries between visible and invisible, and audible and inaudible, occur.

Spatial voice transmission is designed to provide an intuitive way to manage conversation contexts: users can space their avatars in order to control who hears whom. However the
distance over which voice travels is often opaque to users (Erickson et al., 2011). Even worse, the virtual camera in Second Life makes judging audibility impossible (Irani, 2008). A user can be unknowingly overheard in both physical and virtual contexts. During conversation, users need to continually calculate the extent of multiple auras in order to understand whether and by whom they are being overheard, or in general, who can hear whom. In either context people move in and out of one’s aura. In the physical context a phone may ring or someone may enter the room. This is a complex situation to manage and restricts the utility of voice.

8.3.7 Technical difficulties

Setup and sound quality glitches seem like mundane problems, unworthy of analysis. Yet these were reported as significant by voice users in all of my studies. Not all of these are implementation-dependent: voice has inherent complexities relative to text.

The path of sound from sender’s mouth to receiver’s ear involves more steps than the path of text from sender’s fingers to receiver’s screen. Text is already digital at the point it is typed, and need merely be transmitted unchanged through the network. Users can count on their collaborators being equipped with a screen and keyboard that handles text.

By contrast, voice is an analogue signal that must first reach a microphone. Immediately there can be problems: some users have no microphone; those that do may not position it well, or may use a poor quality microphone such as those built into laptops. Headsets are typically fragile, the parts and wires are easily broken and they are more complex to plug in than are keyboards. Many users in my studies had trouble setting microphones up, or complained that their collaborators could not set theirs up.

The distance from mouth to microphone can change unintentionally. Users vary speech from shout to whisper. Microphones pick up background noise, especially when the user is speaking quietly. Voices must be digitized, with a quality that varies from computer to computer. Digital sound consumes vastly greater bandwidth than does text, so that a slow computer or network can easily introduce lag into the transmission. At the same time, lag is more likely to create problems in turn-taking and comprehension when speaking than when typing. For a voice conversation to sound natural, all the voices one is hearing should be at about the same volume, except where change is intentional or due to distance. But the variables just described make volumes differ for no apparent reason. There is no analogy to volume in a text channel.
These complexities mean that the larger the group, the more likely one member will be unable
to use voice, forcing the group to fall back to text or split into voice and text sub-groups. The
systems I studied offered users no feedback on whether they were transmitting too loudly, too
softly, or transmitting background noise. The SL voice client demands extra computing
resources and may be unusable on older machines. For institutional users, I.T. departments
have to be convinced to unblock the required IP ports on firewalls.

The text channel, by contrast, is perceived as reliable and accessible and is used as a ‘safety
net’ when voice is failing.

Technical issues of this kind are rarely highlighted in VW research. An exception is Nilsson et
al. (2002), who studied a series of voice-enabled meetings in ActiveWorlds. My studies
demonstrated that technical problems, while mundane, are a significant part of the experience
of voice users in virtual worlds.

### 8.3.8 Avatars not always necessary

The use of avatars is almost ubiquitous in modern collaborative virtual environments. However
the findings of study 3 suggest that we question whether, and in what circumstances, avatars
are essential to virtual world use.

To ask this question it is necessary first to ask what it is that avatars do. Fundamentally an
avatar is a token that marks a location in the virtual world with a representation of a character
that is situated at that point. The character is either controlled by a user of the system, or is a
non-player character controlled by software. Avatars afford several functions, different
combinations of which are more or less useful in different situations.

Avatars can:

1. Help convey the social presence of users, by adding a visual aspect to the persona that
   users project.

2. Help convey awareness of users’ actions, through visual cues.

3. Afford the dereference of deixis: Related to point 2, the location and orientation of an
   avatar can be a cue to what its user is viewing on screen. This allows fellow users to
   interpret the user’s deictic utterances.

4. Provide a target: In combat games, users either direct fire at each other’s avatars, or
   cooperate to attack the avatars of non-player characters. Note that the target need not
be precisely identical to the avatar – in World of Warcraft for example, experienced users are aware of the ‘hit box’ of a character.

5. Be a prop: In fictional settings, the appearance of avatars helps to set the scene and stimulate user imagination.

6. Be a doll: Some virtual worlds offer customizable avatars, and some users enjoy changing their avatar’s shape, clothes, hairstyle etc.

7. Be a trophy: Users can through their avatars display their achievements to each other.

This analysis allows us to consider in which circumstances the functions of an avatar are required. For simulated combat it is necessary that players provide targets at clearly defined locations. Social worlds require mediated social presence. CSCW requires awareness of collaborators and their actions and the field of work. However my studies indicate that awareness and social presence are mediated less through avatars than through conversation.

**Avatars for coordination of work**

One of the first pieces of research to test coordination through avatars was the ‘furniture world’ study of Hindmarsh et al. (2000), which was discussed in chapters 2 and 6.

Hindmarsh and colleagues studied small groups of people performing joint tasks in a CVE. The study was inspired by the expectation that virtual environments would eventually be used for CSCW:

> In future years, these virtual settings and scenes could well become everyday work or meeting places for remote participants—for example, architects discussing possible alterations to a design; or medical experts discussing and planning surgical techniques (p. 478).

The key problem investigated by Hindmarsh and colleagues was whether users could interpret each other’s deictic references. They were concerned that users:

> could not determine what a coparticipant was referring to, and, more specifically, where, and at what, they were looking or pointing. (p. 480)

Since coordination had already proven to be difficult in media-space research, Hindmarsh and colleagues investigated whether the situation could be improved by allowing users to interact via humanoid avatars that had the ability to point and orient to objects:

> The use of embodiments (or avatars) located in the virtual world provides the participants with access both to the other, and to the other’s actions and orientations in the “local environment.” The embodiments can look at and refer to things and, thus, can be seen alongside the objects at which they are looking and pointing. In this way, and unlike media spaces, (representations of) participants are visibly “embodied in,” and “connected to,” the common world of objects (p. 481).
However this study found interaction through avatars to be problematic. Users often could not interpret their collaborators’ gestures. The narrow field of view meant that the objects being worked on, and the avatars of collaborators, were often not on screen at the same time. Users repaired interaction problems caused by these limitations by keeping up a constant stream of coordination talk.

I found that similar problems occur in a modern virtual world with an avatar system quite advanced relative to the Hindmarsh study. The ‘virtual camera’ in study 3 allowed users to detach from their avatars, and either ‘park’ them or use them merely as place-markers. Schroeder et al. (2006) encountered similar (non-) use of avatars when users played a murder-mystery game in an immersive-projection CVE. While users on meeting each other noticed and commented on each other’s avatars, during collaboration and when the focus was on objects people stopped noticing avatars except as points of reference (p. 663).

Moore et al. (2007) analysed interactions between MMOG players to discover how well avatar systems approximated face-to-face interaction. They asked whether avatars successfully provided their users with awareness of each other’s actions, finding that “avatars display much less information to other players than real bodies do”. They found that avatar eye-gaze did not correspond with what users could see in most MMOGs, and that gestures were difficult to perform (requiring use of menus) and could not be ‘held’ for coordination purposes, but instead ran for a fixed period of time.

There is evidence also from shared-video research that users coordinate primarily through conversation rather than by examining each other’s mediated gestures. In Kraut et al.’s (2002) studies (reviewed in chapter 6), shared video of the work scene but not of head-mounted views improved performance. In Goebbels et al.’s (2003) study, users spent more time looking at the object than each other, and voice quality was more critical than video quality. This indicates that it is more useful to visualize the field of work and where collaborators attention is focused than to visualize collaborators’ bodies.

I expected Second Life’s virtual camera to make deictic reference difficult, perhaps even impossible, and at times this was the case. However users discovered that the camera afforded a new way to use their avatar. While maintaining an (invisible) vantage-point that was optimal for their work, they used their (visible) avatar as a location-marking device. They placed it beside objects and spoke of ‘the place where my avatar is standing’.

Hindmarsh and colleagues proposed several solutions to the coordination problems introduced by avatars:
a. increase users’ field of view through head-mounted displays or peripheral lenses,

b. highlight selection of objects by changing the objects’ appearance,

c. make users’ view-frustrums visible,

d. improve avatar usability, especially regarding control of gestures.

A significant body of work has explored the last of these suggestions, delivering advanced pointing (e.g., Wong and Gutwin, 2010) and eye-gaze (e.g., Murray et al., 2009) and automated tracking of user movements (e.g. Bailensen et al., 2006).

However I would argue that the furniture world study along with my work shows that avatars do not significantly enhance the ability of users to coordinate, and that attempting to fix the problems that avatars introduce may be less efficient than simply not using them in situations where they are not essential.

**Avatars as conveyors of social presence**

It is sometimes asserted that avatars are important for conveying the social presence of users (e.g. Boellstorff, 2008, chapter 5). But the lesson from a succession of post-MRT studies (e.g. Walther, 1996; Nardi et al., 2000; Vetere et al., 2009) is that in contexts both instrumental and recreational, people usually do not need visual representations of their collaborators in order to sense their presence and communicate effectively with them.

Similarly, while it is sometimes asserted that users identify with their avatars, research has shown that identification is likely to be brief and situation-dependent (see discussion in Klimmt et al., 2009; pp. 360-363). The MMOG *EVE Online* works successfully without embodying users as avatars. *EVE* users have a named character but not a public avatar. Only the ship that the character inhabits is visible to other users (Carter, 2011).

Blinka (2008), studying the use of opposite-gender avatars, observed that: “the avatar … tends to be used as a mere game tool without any influential reflection”. In study 2, when gamers continued to speak to collaborators whose on-screen representations had ‘died’, they made it clear that they distinguished visual representation from social presence. My research suggests that it is through conversation as much as avatars that players become present to and aware of each other.

In *Second Life*, where avatars can be heavily customized, some users identify strongly with them. Avatar customization is a popular activity in this virtual world (Ducheneaut et al., 2009).
However in this and other systems, users typically maintain a collection of avatars (Ducheneaut et al., 2004), which limits their identification with any particular one and their ability to project a coherent social presence through avatars alone.

**Avatars as targets**

Some virtual world scenarios, such as combat, require that users perform actions on the virtual bodies of non-player characters or other users. Here the relative locations of characters are a critical part of game state and need to be represented to users.

Most games perform this function through avatars. Users pilot these around virtual space, trying to come close enough to attack the avatars of enemies without being attacked themselves. Raiding requires that players position themselves carefully and have high situational awareness. Even when healers increase their efficiency by displaying charts which obscure much of the virtual scene (Golub, 2010), the relative locations of healer, team-mates and enemy are still important.

MMOG avatars are therefore ‘targets’. They are also ‘props’ in that their appearance supports the game’s back-story, and some are ‘dolls’ in that they afford a degree of customization. But in MMOGs, social presence and a good deal of mutual awareness is conveyed through conversation (Reeves et al. 2009 and my studies 1 and 2). This is why voice makes such a significant difference to sociability and coordination in MMOGs.

Further clues to the function of avatars can be gleaned by comparing computer combat games with text-based MUDs and tabletop strategy and role-playing games. MUDs have no graphics and no avatars, yet cooperative and competitive social interaction still takes place. Users type commands to find out who or what is nearby and type other commands to attack targets thus found. Social presence is conveyed through textual conversation (Kendall, 2002).

In tabletop games such as Warhammer 40k, players are physically co-located and require no mediated social presence or coordination. They control fictional armies represented by miniature soldiers placed on a table-sized game board that represents a fictional or historical setting. These miniatures are both pieces (tokens representing game state) and props (which help to set the scene).

In tabletop role-playing games, players are physically co-present and do not require mediated social presence or coordination. Regardless, some players liked to represent their characters in the form of small model figurines. But these are merely props and take no part in the
calculation of game play. In a recent version of D’n’D, battles are resolved using pieces that represent characters, placed on a representation of the game world. This temporarily transforms D’n’D into a board game, where distance between pieces influences the success of attacks.

**Implications for design**

It is essential that every user of a virtual environment be associated with a virtual location: this is the point from which their view is rendered. Users project a persona into the environment which may be identical to their offline identity or may be a character in a fantasy universe. The question for designers is: In which circumstances and in what fashion should this location or persona be made visible to others?

I listed several functions, the performance of which motivates the inclusion of avatars by designers, and then explored the necessity of these functions in different use scenarios, and whether avatars actually fulfil these functions. My analysis suggests the following design guidelines:

**In MMOGs**, avatars are essential during raids because they perform the tasks of visualizing character locations and providing targets. Appropriately shaped and dressed, avatars are also props that help set the fictitious scene of an MMOG. However user identification with avatars is not essential, nor is the use of avatars to convey social presence, and thus non-anthropomorphic representations such as (temporary) vehicles are fine if this suits the backstory. The mechanics of quests, like raids, are based on proximity to non-player characters: here the non-player characters at least need to be embodied.

Combat is not the only activity in MMOGs. Users engage in discussions, such as for planning before and loot-distribution after raids. During discussion users can enact proxemic relations with each other’s avatars in order to control membership of conversation groups (Benford et al., 1994; Yee et al., 2007). But as we have seen (studies one and four, and Erickson et al., 2011), using avatar placement to control message transmission is fraught with error, and the analogy with offline conversation groups is often misleading. MMOG users usually choose other ways to define the members of a conversation, such as team channels.

**In open-ended worlds**, the necessity to perform actions upon the ‘bodies’ of other users is limited or non-existent: thus avatars are not required to be targets. Avatars may perform other functions but these are situation-dependent. Some users enjoy customizing and dressing their
avatars, while others do not. Avatars with appropriate appearances enhance enjoyment of the fictional setting for some users.

Social presence is important in some situations, and this can be enhanced by avatars, though well-designed linguistic communication tools are probably more important. Group activities occur in social worlds – discussions and meetings are a popular example – and avatars can assist with mutual awareness, though this works far from perfectly and many users forget about avatars during these activities. Many Second Life users create objects or admire the creations of others - where this is performed by groups, deictic reference can be useful, but this is not supported well by avatars, and representing reference directly on the object may work better.

**In systems designed to support computer-supported cooperative work**, the ability of avatars to be targets, props or dolls is unimportant. Social presence is important, though in a workplace system this is likely to be effected through talk. For object-focused work, awareness and the ability to use and understand spatial deixis are critical; however study 3 and earlier research shows that avatars have limited utility in this regard. During workplace collaboration it is conversation, rather than the mutual viewing of avatars, that best supports social interaction. In chapter 6 I proposed alternative mechanisms for supporting users’ talk about locations and objects in virtual space.

At IBM, where virtual worlds were championed as workplace tools, executives complained that avatars were difficult to use, gave misleading impressions as to who was in a conversation group, gave misleading impressions as to the identity of their owners and were hard to recognize (Erickson et al., 2011). My own experience using Second Life for meetings and teaching suggests that people quickly forget about their avatars once conversation starts. It would be ironic if, in the attempt to broaden the uses to which virtual worlds are put, designers were to over-emphasize the necessity for embodiment in avatars, in a way that reduced the usability of virtual worlds and repelled potential users.
8.3.9 Instrumental use of virtual worlds

I end this section by offering a view on the potential for non-recreational use of virtual worlds. Virtual environments have been promoted as tools for commerce, education and government (e.g., Bessiere et al., 2009). Yet uptake has been disappointing in these domains (although Wasko et al., 2011, claimed that virtual worlds have emerged from their ‘trough of disillusionment’). At the same time the ‘2d’ web has been very successful in all of these domains and as a recreational tool. This disappointing uptake despite significant promotional effort by vendors and enthusiasts prompts the question: To which domains are VW technologies suited? Or equivalently, which properties of virtual worlds are impeding their uptake beyond recreational domains?

Schroeder (2010) identified two technologies by which users can achieve the sense of ‘being there together’: collaborative virtual environments and video-conferencing technologies. Both have been plagued with practical problems of cost and capacity, though Schroeder envisaged ideal technological ‘end points’ in order to analyse to what purposes these technologies might be put when they have matured.

Both videoconferencing systems and CVEs provide a visual image of one’s collaborators in space, but they differ in how the image is produced. Video-conferencing transmits images of physical bodies situated in physical space. Virtual environments show computer-generated bodies in a computer-generated environment. Post-MRT research has suggested that sub-video-fidelity representation of collaborators is usually sufficient and frequently desirable, so the lack of photo-realism in CVEs should not be expected to be a problem. Yet CVEs have been resisted by most business users.

So what are the situations in which it is desirable to ‘be there together’ in a computer-generated space? Video-gaming is one obvious answer: a large part of the appeal of multiplayer games lies in seeing oneself and one’s collaborators as characters in a fantasy environment. But are there instrumental tasks to which a computer-generated world is suited?

My experience suggests that virtual worlds are most likely to be preferred over the web in scenarios when they allow their users:

(a) to experience a simulation of an activity or place, which is

(b) useful, interesting or enjoyable to the user,

(c) difficult, dangerous or expensive to experience in an unmediated fashion, and
(d) where three dimensional space is essential to the experience.

Analyzed according to these criteria, vehicle simulators are successful because they afford an educational or entertaining experience which is dangerous or expensive in the physical world. Combat games simulate activities which are highly spatial, fascinating to many, yet horrific to experience in reality. MMOGs add to combat the ‘virtual work’ of teams and a career structure, which some enjoy, in a large space which can be explored and which offers social interaction with people both known and unknown.

Using the same criteria we can predict that simulations of activities which are mundane, cheap or annoying, or to which spatiality is not core but merely an inconvenience, will not succeed. The belief that 3d is either necessary or sufficient to make a communication technology attractive lies behind several market failures (Yee et al., 2009). Some technologies failed because they reintroduced to online communication the need to negotiate space: a requirement that earlier telecommunication technologies were designed to ease. Most people do not want to navigate an avatar through virtual space in order to attend a meeting or obtain information from a virtual kiosk: they use the Internet to avoid having to do that.

Nah et al. (2011) found that while 3d environments induced a greater sense of flow and telepresence than did 2d environments, they had a negative impact on brand equity.

On the one hand, the 3DVW environment is immersive, fun, and interesting, which can enhance users’ learning and strengthen the impact of their exposure to the brand. On the other hand, having to navigate and interact in a 3DVW site can be a distraction to the subjects in attending to the audio and visual information on the site. (Nah et al., 2011, p. 742).

The four criteria I proposed above suggest that virtual worlds will be adopted for tasks which are difficult, expensive or dangerous to perform without the VW, and inherently spatial. This suggests use-cases such as tele-medicine, disaster recovery and remote repair. In these situations workers on-site deal with physical objects, with assistance from helpers who are distant to them. While video will clearly be useful in many of these situations, computer-generated objects might have advantages too, such as support for planning, the ability to visualize hypothetical manipulations, and the ability to “explode” a complex object to view relationships between the parts. PC-based virtual worlds will have an advantage over larger, more unwieldy immersive technologies in that they can be accessed via Internet-enabled laptops which can be carried into the field.

At the same time, my results suggest that of the component technologies that comprise a virtual world, avatars are the least important during workplace use. And if avatars are used,
having them mimic their owners’ gestures with high fidelity will be less useful than making virtual objects themselves accountable and providing users with highly usable marker objects.

The disadvantages of voice with regards to over-projection of offline identity will not manifest in most work scenarios. Studies 1 and 4 suggest that either linguistic modality can convey sufficient social presence for work to be conducted. Ideally both would be available, to be chosen as the situation demands.

### 8.4 Summary of contributions

In summary, the contributions of the thesis are.

1. I demonstrated how phenomena already identified in virtual worlds research, such as immersion, identity play and gender play, interact with communication modality.
2. I demonstrated how phenomena already identified in communication research, such as context, group size, impression management and eavesdropping play out in virtual worlds.
3. I demonstrated the mapping from uses and user types to preferences for communication modality.
4. I identified use scenarios that are preferentially suited to either the voice or the text modality.
5. I confirmed that in the virtual worlds context, media richness (i.e., high-fidelity projection of the user’s social presence) can present both problems and advantages.
6. I explored the impact of integrating voice channels with the spatiality of a VW.
7. I updated our understanding of problematic spatial reference in the context of modern VWs.
8. I showed that VW users exploit the leanness of media in order to multi-task, and argued that this calls into question the value of immersion in the VW.
9. I analysed the functions that avatars perform, and discussed their suitability to various VW scenarios.
10. I showed that users discard the avatar in some situations, and/or use it for purposes other than were intended, and called into question the necessity for embodiment in avatars.
8.5 Limitations of my research

In this section I address questions of the generalizability of my findings, the shortcomings of my methods and the limits of my data.

I have followed the advice of Glaser and Strauss (1967, pp. 228-230) and attempted to convey the credibility of my findings by presenting together both emerging theory and examples of the data in which it is grounded. In appendices C to F I have included details on how the data were gathered and analysed: this is illustrated through research documents, sample data and coding hierarchies. Likely sources of error, such as sampling bias, are discussed within the study chapters.

There are at least five major sources of variation in the domain I studied:

- virtual worlds
- voice system configurations
- users
- use-cases, and
- contexts both virtual and physical.

It is not possible in four studies to adequately sample the available variation along each of these dimensions. My strategy has been to seek diversity where I could, both within and between studies.

- I examined both major types of VWs – social and game worlds.
- I examined both major voice configurations: radio and spatial.
- I observed groups both known and unknown to each other.
- I recruited participants with a variety of approaches to using VWs.

One way of extending this research would be to address other sources of variation. For example, voice is now available in new virtual worlds not examined here. Some of these differ in interesting ways from the systems I examined, and it would be valuable to study them.

All of my participants were adults aged under 50, and many of them lived in Australia. It would be interesting to test whether my results hold across other demographic categories.
8.6 Possibilities for future research

Second Life recently introduced a Voice Morphing feature which allows users to change their natural voices using a menu of theme effects. The promotional video for this feature promises: “You can match your voice to your avatar”, and “You can be a whole new persona”. It would appear that this technology is designed to address the dilemma I have described in this thesis: that while voice is a natural, efficient way to communicate, it conveys more social information than many users desire, and interferes with identity-play. It would be interesting to see whether voice-morphing solves the problems of voice in virtual worlds without introducing new ones. I note that Xbox Live offered a similar feature, but that our earlier study discovered that most users rejected it (Gibbs et al., 2004).

Avatar Kinect is a new system incorporating virtual space which uses Microsoft’s Kinect device to make an avatar mimic the movements of its user. While motion-tracking has been used experimentally to control avatars (e.g. Bailenson et al., 2006), this is the first widely-available commercial product that offers the feature. It will be interesting to study, in a commercially available system, the use of motion-tracking to control avatars, and to see whether it is adopted and what benefits and problems it brings. I predict in section 8.3.3 that gesture-tracking may, as voice did, create problems involving the over-transmission of social information into virtual space.

Avatar Kinect has been demonstrated as a tool for online discussions. In these scenarios avatars sit in a group and talk, reflecting their users’ actions who of course doing the same thing. We can extrapolate from the controversy over voice to predict that the usefulness of gesture-tracking will be limited in scenarios other than meetings. Many users don’t want their in-world presence to mirror their offline reality too closely.

Gesture-tracking could improve guild meetings in World of Warcraft, or make discussions in Second Life more engaging. But even augmentationists attending meetings are unlikely to want their every offline activity projected into the virtual world. It is likely that a ‘mute’ button will be required, as it is in microphones, and this will add complexity to the UI.

---

14 see http://secondlife.com/landing/voicemorphing, accessed on 1st December 2011
8.7 Conclusion

I conducted four studies of voice communication in virtual worlds, finding that the introduction of voice transforms the user experience. My work shows that tension over media richness persists in the VW domain, and that mundane problems of sound quality and reliability have a significant influence on user experience.

My work emphasizes that virtual worlds are a unique context for communication. Complexity arises out of users’ desire for control over projection of offline identity, and their ability to conduct multiple simultaneous conversations – and be overheard – in both physical and virtual locations, each with its own physics of visibility and audibility. A voice-enabled virtual world can become a maelstrom of impression management, identity play, and confusion over what is being transmitted to whom. Prior research into mediated communication sheds some light on the problem, but studies such as the ones reported here are necessary in order to capture the rich influence of voice on virtual worlds.
9. Bibliography


Appendix A: Typographical conventions

When quoting participants or authors inline, I have used double-quotation marks, for example: “I really enjoyed using this technology”.

I separated longer quotes from the surrounding text, setting them in a smaller font, indented and single-spaced. For example:

It just seems a lot more natural being able to talk to people and say “come over here, check this out”, as opposed to having to type stuff.

The distinction between inline and indented quotes has been made for readability purposes and doesn’t indicate a difference in the source, nature or importance of the quote.

Quotes from prior publications include page numbers except where these are not available such as for online publications and some conference papers.

Sometimes, for clarity, I have included an interview question in the transcript. These are italicized. For example:

They can hear you fine.

Yeah, and some people are crystal clear.

Occasionally I have had to edit interview transcripts for clarity, for example replacing pronouns with their antecedents. In these cases I have placed my edit in square brackets. For example:

I think that [voice] broadens, but reduces the depth of the communication.

Where I have omitted part of a quote, I mark this with an ellipsis in square brackets, e.g.

… those who want to take something other than RL there. […] In the early stages of Second Life …

To mark colloquial expressions such as ‘pickup group’ I have used single-quotes.

Names of commercial systems are in italics, e.g. World of Warcraft.

I have sanitized some offensive words using italics, e.g. f***. My aim is to avoid offending sensitive readers, while at the same time allowing readers who prefer ethnographic realism to decipher what was said.
### Appendix B: Glossary

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Camera</td>
<td>the point in virtual space from where a user’s view is calculated</td>
</tr>
<tr>
<td>CSCW</td>
<td>Computer-Supported Cooperative Work</td>
</tr>
<tr>
<td>CVE</td>
<td>Collaborative Virtual Environment</td>
</tr>
<tr>
<td>DDO</td>
<td><em>Dungeons and Dragons Online</em></td>
</tr>
<tr>
<td>F2F</td>
<td>face-to-face</td>
</tr>
<tr>
<td>ICT</td>
<td>information and communication technology</td>
</tr>
<tr>
<td>IM</td>
<td>instant messaging</td>
</tr>
<tr>
<td>I/O</td>
<td>input/output</td>
</tr>
<tr>
<td>in-avatar</td>
<td>the state of having attached one’s view-camera to one’s avatar</td>
</tr>
<tr>
<td>in-camera</td>
<td>the state of having detached one’s view-camera from one’s avatar</td>
</tr>
<tr>
<td>in-world</td>
<td>In the virtual world (expression used by <em>Second Life</em> users)</td>
</tr>
<tr>
<td>MMOG</td>
<td>massively multiplayer online game</td>
</tr>
<tr>
<td>MMORPG</td>
<td>massively multiplayer online role-playing game</td>
</tr>
<tr>
<td>Prim</td>
<td>primitive shape – the building blocks of <em>Second Life</em></td>
</tr>
<tr>
<td>PUG</td>
<td>Pickup group (temporary grouping of MMOG players)</td>
</tr>
<tr>
<td>Roger-Wilco</td>
<td>voice-over-IP system intended for use by video-gamers</td>
</tr>
<tr>
<td>Vent</td>
<td>Ventrilo, voice-over-IP system intended for use by video-gamers</td>
</tr>
<tr>
<td>TS</td>
<td>TeamSpeak, voice-over-IP system intended for use by video-gamers</td>
</tr>
<tr>
<td>SL</td>
<td><em>Second Life</em></td>
</tr>
<tr>
<td>UI</td>
<td>user interface</td>
</tr>
<tr>
<td>VoIP</td>
<td>voice over IP (IP = ‘internet protocol’)</td>
</tr>
<tr>
<td>VW</td>
<td>Virtual World</td>
</tr>
<tr>
<td>WoW</td>
<td><em>World of Warcraft</em></td>
</tr>
</tbody>
</table>
Appendix C: Publications arising from this research

Conference presentations (in date order)


Workshop presentations


Appendices

Book chapter

Appendix D: Material from study 1

This appendix contains excerpts of the ethics, data and coding documents from my first study on voice in MMOGs.

D.1 Human Research Ethics Committee documents

First I present the plain language statements (PLS) that were approved by the University of Melbourne’s Human Research Ethics Committee. These were to be read by study participants. There were two PLS: one for the participants who were assigned to play DDO, and one for the participants who played their existing MMOG.
THE UNIVERSITY OF MELBOURNE
DEPARTMENT OF INFORMATION SYSTEMS

Research Project Description

PROJECT TITLE: Voice Communication in Massively Multiplayer Online Games (Phase I)

INVESTIGATORS:
- Mr. Greg Wadley
  [greg.wadley@unimelb.edu.au]
- Dr. Martin R. Gibbs
  [martin.gibbs@unimelb.edu.au]
- Ms. Pam Binda
  [pbinda@unimelb.edu.au]

The aim of this project is to examine the use of voice communication technology in Massively Multiplayer Online Games (MMOGs).

Voice technology for player-to-player communication is standard on the Xbox Live game network, and is becoming a popular choice in PC-based multiplayer games as well. However, while voice may be superior to text for coordination of teams in fast-paced games, it is not known whether voice is suitable for MMOGs, where many players must communicate over long periods within large persistent virtual worlds. This project will determine factors important to the design of voice communication in MMOGs.

We would like to interview you about your experiences in MMOGs and your experience with using text and voice to communicate with other players in these games. The interview will take up to 90 minutes of your time. It will be video recorded for later analysis by the named researchers.

The project does not involve any risks. The only anticipated inconvenience for you is the time taken by the observation session and focus group. The information you provide will be treated as confidential and used for research purposes consistent with this research project only. Confidentiality of the information provided will be protected, subject to any legal limitations. Access to the information will be restricted to the investigators only. As required by the University, data will be held in locked cabinets in the Department of Information Systems, and destroyed using confidential waste disposal techniques after five years following last publication from the research. No individual person or organization will be identifiable in the research report written up about any gaming session and focus group. However, due to the small number of participants in this research project, even though people will not be named, there is a possibility that people could be identified by contextual information.

Your participation in the research project is voluntary, and you may withdraw your consent to participate and discontinue participation at any time without prejudice. You may also withdraw any unprocessed data previously provided.

The project is supported financially by Microsoft Research Asia. However only the named researchers will be directly involved in the research.

If you have any questions about this research project please contact Greg Wadley on 8344-1386 or greg.wadley@unimelb.edu.au, or Martin Gibbs on 8344-1394 or martin.gibbs@unimelb.edu.au.

If you have any concerns regarding the conduct of the research please contact the Executive Office, Human Research Ethics, The University of Melbourne 3010, ph: 9344 7207 fax 9344 6739.

Figure 21: PLS 1 for study 1
Next are the interview questions and diary instructions for participants. These were also approved by the University Human Research Ethics Committee.
Draft list of interview questions

Both phase one and phase two of our research involve individual or group interviews. These will be conducted using an open, semi-structured interview process. We will use the following list of questions to stimulate discussion, however discussion topics will also emerge from observation of play, and during the interview itself.

- Did you like using voice-over-IP in the game?
- Was it better or worse than using text?
- How often did you use voice compared to text?
- Did your use of voice change during the three month study?
- Were their particular game situations in which either voice or text was preferred?
- How well was the voice communication implemented?
- How could voice be better implemented in MMOGs?
- Were any aspects of the user interface for voice frustrating?
- Do you prefer a ‘always on’ or a ‘press to talk’ interface for voice?
- How much have you played MMOGs before?
- Have you used voice in games before? In what kinds of games? Was it useful?
- Is voice more useful in fast team games such as Counterstrike than in MMOGs?
- Is voice better for small or large groups of players?
- Is voice better for person-to-person, vicinity, quest, or entire-clan chat?
- Is voice better for some styles of play than others?
- Is voice better in some parts of game play than others?
- What communication technology would you like to see introduced in future MMOGs?

Figure 23: Interview questions for study 1
Next I present sample data from participant interviews.
D.2 Excerpt from interview with [anonymized participant A]

[25:15 – talking about WoW] You need some way to know who’s talking. There is some channel control functionality. My guild uses Ventrilo, and I think that’s got the ability for the officers in the Ventrilo server to mute people. That does happen on occasion, when we’re in large-scale raids, like there’s 40 man instance, and someone won’t keep their mouth shut, it’s getting in the way of the business at hand, they can get muted. You need that.

[26:0] It’s quite common for the officers to go to the officer’s channel, while they discuss issues, and then they’ll drop back into whatever the raid channel is.

[26:55] It’s a very obvious observation that one of the benefits of text over voice is that people talk over each other, like in real life. In text you can’t do that. In text chat you can see everything that’s going on even if you don’t have time to keep with it simultaneously; you can refer back to what was said. There’s a history. You can’t have multiple threads in voice: it’s quite dysfunctional. I find in my guild, in which it’s compulsory to use Ventrilo when you’re doing these raids, they still use text to convey essential messages, so that there’s no possibility that the message is not being conveyed. Because the raid leaders and officers – if you make a really stupid mistake, and they get really crapped off, and for good reason – you might spend three hours on a Saturday afternoon trying to get to an endpoint, and If someone f***s it up because they’ve done something stupid, so they make it quite clear in text that these are the dos and the do nots, as well as chatting about what they’re doing in the voice client.

D.3 Excerpt from interview with [anonymized participant B]

[1:00] I joined *World of Warcraft*, I was just playing by myself, questing with a few people, and then I joined a guild which had a Ventrilo, and now I’m questing with them and doing instances with them, and talking through the voice at the same time. There’s a leader and they’ll tell you to do the things. Voice is useful for doing raids, versus other real people, and we use it for just chatting about stuff we’ve done in the daytime, and having a joke, that sort of stuff.

[3:44] In the guild, not all of them use Ventrilo, there’s about 70 people in the guild now. They’re not all online at the same time. There’s usually 20 to 30 people online in the guild. It’s quite an active guild. There are different channels. There’s usually only about 10 people on
Ventrilo because not everyone uses it – some people don’t like it. Some people would rather not talk to people. It’s usually fine. People take turns.

[4:45] We have a raid channel, and an officer channel for the officers to talk. So if we go on a raid we’ll get moved into the raid channel, and there’s general chat, and other stuff.

[5:35] On a raid, they won’t let you into the raid unless you’re on Ventrilo. They say “use it or don’t come”.

D.4 Excerpt from interview with [anonymized participant C]

[2:37] Voice is a lot better than text. Myself and another guy we played DDO with – a couple of times we jumped into WoW and we found it very awkward not to have the voice at our fingertips. Using text chat was ok – it just seemed a bit lacking. I wouldn’t have realized before using voice just how much of a difference it makes to playing the game.

When I was first playing DDO I did some pickup groups. There were a couple of times I teamed with [name], we had people join us and stuff, but especially in the early days we found only one or two people in a pickup group using voice. It didn’t seem to get used that heavily. It seemed to be mainly [name] and I who were doing all the talking. I don’t know if that was because we knew each other. We were just having general fun banter, role-playing, and doing the dungeons. You might occasionally get something but usually we had to direct a question to a person to get a response out of them. I don’t know if they just weren’t relaxed to talk.

[4:20] There were some people who didn’t have voice at first. We got them to enable voice so they could at least hear us talk, but they would have to type to ask us. There was a mix of us talking and typing as well.

I don’t think our use of voice changed during the study.

[5:35] Once we got over the initial teething problems they were having with voice, we found it was very much an integral part of the game. This was strange because before I was playing DDO I was playing on a role-playing servers in WoW, and they were talking a bit in voice there, and myself and a couple of others had actually been against voice because we thought it would detract from the role-playing, we found it was easier to role-play using text. Myself and a couple of the other guys were playing female characters, and we thought that it would detract from the immersion. The thing that surprised both of us was that some of the females didn’t realize that we were male until we met – we were talking about it, they were mainly US
people and they said “What, you’re a guy?”, It didn’t change the way we played after that when they realized we were guys.

D.5 Excerpt from focus group interview

[This is the group whose members knew each other prior to the study.]

[11:15] You’ve got that ladder that goes down [in one of the DDO dungeons]. So we’re trying to direct him [a team-mate] via voice. He’s almost on the ladder and then he goes off to the side. We go “no no no stop stop!” and he kept on walking, and then I’m “ok forward forward forward”, and then we’ve got him lined up, and then he started to back up, and for some reason he went off-centre a bit, and so he’s backing up and backup up and we’re screaming “stop! stop!”. It must have been because there was enough lag there, he didn’t hear it, so he just went [falling noise] straight to the bottom of the pit.

[11:30] There was one I think where [name] was playing classical music in the background, and every so often we’d suddenly get the classical music, the peaks of the music, coming across [the voice channel].

Yeah, I’d activate the voice and it would pick up [the music] over my voice.

And your kids in the background. Everyone now and then I’d hear these blood-curdling screams. [laughter]

[12:20] Also I can’t use voice while my wife’s watching TV.

[13:20] Crunch crunch crunch – are you enjoying your biscuits [name]?

[13:34] I think with us, because we’ve known each other for a fair while, it makes a difference with regards to funny things, because we’re all familiar with each other and so we joke around anyway. And the flip side, the agro side not the humour side, that never seems to come up. That familiarity makes a huge difference I think. And it always makes a difference half way down a bottle of wine.

[15:40] Have you guys every used voice with people you don’t know?

[15:00] One group that really tripped me out, they were very sparse. They had voice on, but none of them spoke unless, and they were almost monosyllabic, they’d done ‘Water-works’ so many times, that they’d simply say “do this, do that”. You heard nothing else. They just did the run, gave the cues, that was it. I tried a couple of times to interject some conversation into it – there was just deathly silence.
I’ll always scream “trap trap trap!” [laughter]. I always thought it was an ice-breaker but apparently not.

D.5 First pass at code hierarchy for study 1

Immediately after my data gathering I analyzed the data into a hierarchy of themes, which was used for a conference publication (Wadley et al., 2007). This hierarchy is illustrated in figure 26. While writing this thesis I re-analyzed the data to arrive at the new hierarchy shown in figure 27.

- **Game-play and grouping in DDO**
- **Appropriating voice in DDO**
- **Who says what to whom?**
- **Benefits**
  - voice is more natural
  - voice usefulness depends on game activity
- **Problems**
  - technical problems with voice hardware and software
  - problems with use of voice
  - not thinking before talking
  - voice is synchronous only
  - voice is harder in large groups
  - real world leaks into the shared audio space
  - does voice affect role-playing?
- **Design suggestions by participants**

*Figure 25: First pass at code hierarchy for study 1*
## D.6 Code hierarchy for study 1

This screenshot from NVivo illustrates my final coding hierarchy for study 1.

<table>
<thead>
<tr>
<th>Category</th>
<th>Sources</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PRO</strong></td>
<td>12</td>
<td>35</td>
</tr>
<tr>
<td>doesn't fit the genre</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>text-paced action</td>
<td>12</td>
<td>20</td>
</tr>
<tr>
<td>other players feel present</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>get to know each other better</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>negotiation</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>don't like text</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>CON</strong></td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>don't like voice</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>ruins the magic</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>gender play</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>channel clutter</td>
<td>10</td>
<td>21</td>
</tr>
<tr>
<td>irrelevant profile</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>can't tell who's talking</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>technical problems</td>
<td>11</td>
<td>16</td>
</tr>
<tr>
<td>abuse, gritting</td>
<td>5</td>
<td>14</td>
</tr>
<tr>
<td>shy people uncomfortable with voice</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>offline sound leaks into game</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>extended overhear voice</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>children might hear swearing</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>the voice-text divide</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>how to pronounce game words</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td><strong>both PRO and CON</strong></td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>personal characteristics exposed</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>immediate reaction</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td><strong>USING VOICE</strong></td>
<td>14</td>
<td>100</td>
</tr>
<tr>
<td>what voice is used for</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>popular</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>compulsory</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>socializing</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>multiple channels</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>listening but not speaking</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>role-play</td>
<td>10</td>
<td>18</td>
</tr>
<tr>
<td>real name</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>national accents</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>radio discipline</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>quantity of speech</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>out-of-game communication</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>whether people know each other</td>
<td>11</td>
<td>23</td>
</tr>
<tr>
<td>voice integrated with game</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>using both Text and Voice</td>
<td>8</td>
<td>10</td>
</tr>
</tbody>
</table>

*Figure 26: Code hierarchy for study 1*
Appendix E: Material from study 2

E.1 Human Research Ethics Committee documents

Study 2 was conducted using the ethics application from study 1, with modified instructions for participants as shown in figure 23.

![Instructions for Participants](image)

*Figure 27: Participant instructions for study 2*

E.2 Excerpt from focus group interview

[This group interview took place at the end of the trial.]

[6:15] One thing that struck me [watching a session] was, one of the teams went “oh let’s all wait” and the proximity nature of it got them to move together rather than run off individually.

*Did you find that it’s changed your actual game-play?*

Definitely. Before, we never planned anything, we’d just run off. [laughter]
It would have taken too much energy to type it. [laughter]

But now pretty much before each game we seem to huddle together and decide, ok you’re going to go ahead, you’re going to be medic, and let’s try to have some sort of a plan. Before it was pretty much everyone on their own.

It’s very easy and natural to do that with SV. You can just react to a situation immediately and call out “look out” or whatever, and don’t think twice about it. You can actually have that capability.

Yes I think the nature of the game is completely different with SV and without SV. Like if you play on an online server [not equipped with SV] you have a team, but for all intents and purposes they’re meaningless to you. It’s about your stats and what you’re doing. Whereas with SV it’s a lot more about how the team is performing as opposed to how you as an individual are performing.

Yes ET is particularly a team object game. So [SV] really suits that game as well.

And how would that compare with a normal radio-voice product? Do you think spatialization adds anything?

Yes I think it makes a big difference because you know you can talk just to the people around you. So it’s not like team commands, you’re just talking to everyone, so you can heckle [laughter] … It’s not a tactic thing.

But then if you’re using Ventrilo and you say “I need some help”, you’ve got three medics in the team and they’re saying “where are you?” Whereas if you’ve got spatial audio and you see the medic, there’s only going to be one medic around who can hear you, and they’re: “Oh, you’re the one who needs help, you can have some health packs”. That does help compared to Ventrilo.

I have a slightly different position. I think there’s a very distinct difference with game play and the performance of the individual or a team. I think products like Roger-Wilco and Ventrilo rillo were targeted at people who want to achieve higher performance as a team, because for example you can speak when you’re dead, or you can speak only to your team – the enemy can’t hear you. All these aspects that give you an edge over your opposition. Whereas SV is about adding an extra element altogether to the game.

It’s interesting. I agree. Sometimes you’re in a position where you can hear the enemy. They might be hiding around a corner and you hear them say “throw an air-strike in” or they’re talking about the tactic that they’re going to use. So you can really take advantage of the
knowledge of what they’re going to do. This wouldn’t be there with Roger-Wilco or other systems. So it definitely can be a disadvantage to be using SV, but I think it does add to the game-play.

E.2 Code hierarchy for study 2

![Coding hierarchy from study 2](image)

*Figure 28: Coding hierarchy from study 2*
Appendix F: Material from study 3

This appendix contains excerpts from the ethics, data and coding documents from my third study on collaboration around objects in *Second Life*.

F.1 Application to IRB for Human Subjects Research

In this section I present my application to PARC’s Internal Review Board for human subjects research, followed by the advertisement I used to recruit participants.
Collaborative Virtual Building study

Application to PARC’s Internal Review Board for Human Subjects Research
Prepared by Greg Wadley
6th August 2008

Background and Purpose
The Collaborative Virtual Building study is intended to gather data on how users collaboratively manipulate complex objects in the virtual world Second Life (SL). Collaborative manipulation was a focus of VR and CVE researchers in the 1990s but has not been studied in the popular virtual worlds that have appeared since then. It is hoped that the study will produce design insights for the CVE project.

Study Team Information
The study is being run by Greg Wadley (visiting researcher) under the supervision of Nic Ducheneaut.

Participants
We will recruit 20 to 30 people from the Bay Area using Craigslist and the Second Life user forum. Participants are required to have a reasonable amount of experience with Second Life. They will range in age but will be at least 18 years old. All participants will be asked to read and sign the attached consent form, which informs them about the study, including the types of data that will be collected and how they will be used. It also explains that the building tasks and interviews will be audio recorded and explains how these recordings will be used. All participants will be given the option to refuse or discontinue participation in the study at any time.

Procedure
The study will last one hour per participant. Participants will need to travel to PARC. Groups of participants will sit together in a lab and carry out two building tasks while logged into Second Life. After the tasks have been completed, the researchers will conduct a focus group with the participants to discuss their experience and gather suggestions about how well the Second Life user interface supports group building.

Data: The data for the study consists of:
- Screen video from the Second Life client
- Audio recordings of participants’ voices while using Second Life
- Audio recording of the focus group
- Researcher notes made while observing task and focus group

We will pay each participant $30 for the session.

Confidentiality
All the data about participants will be kept confidential. Each participant will be assigned a code that will be used in all the data analysis. Only members of the research team will have access to information linking names to the assigned codes. Anonymous written transcriptions of audiotapes and short audio segments of the interviews may be used in publications and presentations for the scientific community or for business development activities.

Potential Risks
We are not aware of any risks that this study may entail.

Figure 29: IRB application for study 3
F.2 Sample data

Here I present samples of the data I collected in this study. These includes: notes taken during lab trials, quantitative analysis of trial videos, and an excerpt from a Second Life transcript. I have obscured participant names.
F 2.1 Notes taken during lab trials

I took notes during trials and wrote them up afterwards while my memory was fresh. Here are three samples.

Figure 31: Study 3 lab notes 1

These guys worked well together. They know each other – especially W and R. They all had videogame experience, and some had a little CAD experience. A was an A user, W a B, and R either A or B.

They were the first group to use cardinal directions (north, south, east, west) to orient themselves. A seemed to come up with the idea first – though R is an experienced hiker who confirmed that she likes to orient this way in real life. We talked about how Bay Area residents tend to say “north” or “south” when the roads they are describing actually run NW-SE.

R: “If I’m standing here, in front of me is in front of me, but when I’m moving the camera in front of me is over there.”
“I forgot I was the cat.”

When asked about extra UI to support group work, there were no particular suggestions. I suggested an extra screen for collaborator view and there was a little enthusiasm for that. We talked about being able to sense teammates stress levels and respond to that, which doesn’t actually answer the question but is interesting.
At first these two guys often tried to edit the same wall of the house. They didn’t seem to communicate at all about which object they were going to work on. Overall they didn’t talk much. Amount of talk varies a lot from group to group.

At one point [redacted] said to [redacted] “the one on your right”. But this is not useful if [redacted] has moved his camera. Did he know that he didn’t? At 12:37 [redacted] said “I’m going to put the chimney on your left” and it wasn’t on P’s left as he was alt-zooming. But P saw it anyway and just continued.

At 23:12, [redacted] suggests they go outside the house for a better view of the wall they are fixing. But [redacted] already has his camera outside. P doesn’t know this, and G probably doesn’t know that P is assuming he is inside.

At 24:31 they seem to be talking about different gaps. (These are fixing gaps where roof meets walls.)

In the furniture task, P is saying “bring the table closer to me”, but G is not looking to see where P’s avatar is. Eventually P goes and stands his avatar at the desired spot, masking it for G to see.

C, the experienced SL user, did some things differently. For example he knows that objects can pass through objects. So when he moves the roof onto the house, he brings it up through the house. Notice users (including me) always bring it around the over the top of the house and lower it, as a crane would IRL. The novice user perceives that they are moving objects in a world, and considers physics. The experienced user sees it all as drawing a picture, and moves anything anywhere according to efficiency of drawing.

P said wants a change to UI – a pointer, laser-pointer. C helped the designer of Camecaster promote it. But hasn’t used it.

Figure 32: Study 3 lab notes 2
I engaged in *Second Life* classes, competitions and other building activities and discussed problems that were arising in the study with people I encountered. Here is a sample transcript from one discussion, conducted in text.

---

Figure 33: Study 3 lab notes 3

---

I engaged in *Second Life* classes, competitions and other building activities and discussed problems that were arising in the study with people I encountered. Here is a sample transcript from one discussion, conducted in text.
I analysed the trial videos, noting when participants switched from in-avatar to in-camera view, and when they attempted to refer to an object, classifying the latter into three references types (with respect to Avatar, with respect to Object, and by Description). After annotating the video I calculated for each participant the proportion of time they spent in each camera mode. The following screenshot from Excel shows a portion of data for one lab trial. In columns B through F I recorded events, while in column A I recorded the time of the event. Columns B and C show when the two participants move from one camera mode to the other. Columns D through F record instances of deixis using the first letter of the participant’s name. Names are obscured.
<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Avatar vs Camera</td>
<td>References to Objects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>wrt Av</td>
<td>wrt Ob</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:03:31</td>
<td>start in av</td>
<td>start in av</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:04:08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:02:16</td>
<td>camera</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:02:20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:02:40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:02:49</td>
<td>camera</td>
<td>avatar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:03:01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:03:30</td>
<td>camera</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:03:34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:03:45</td>
<td>avatar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:04:26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:05:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:05:14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>G</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:05:16</td>
<td></td>
<td>avatar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:05:24</td>
<td>camera</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:05:48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:07:10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:08:30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:08:45</td>
<td>G</td>
<td>P</td>
<td>'left, right, back'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:08:12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:09:40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:10:05</td>
<td></td>
<td></td>
<td>G</td>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:10:16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:10:26</td>
<td></td>
<td></td>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td>P</td>
<td>G</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:11:30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:12:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:12:22</td>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:13:30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:13:50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:14:25</td>
<td></td>
<td>avatar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:14:35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:14:42</td>
<td>camera</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:17:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:18:45</td>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:21:14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:21:00</td>
<td>avatar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:21:41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 35: Sample video analysis: one pair
Appendix G: Material from study 4

This appendix contains excerpts from the ethics, data and coding documents from my fourth study on voice in Second Life.

G.1 Human Research Ethics Committee documents

First I present the plain language statement (PLS) that was approved by the University of Melbourne’s Human Research Ethics Committee, to be read by study participants.

---

THE UNIVERSITY OF MELBOURNE
DEPARTMENT OF INFORMATION SYSTEMS

Research Project Description (plain language statement)

PROJECT TITLE:  Using voice to communicate in Second Life (interview group)
INVESTIGATORS:  Mr. Greg Wadley  (greg.wadley@unimelb.edu.au)
                 Dr. Martin Gibbs  (martin.gibbs@unimelb.edu.au)

The aim of this project is to examine the use of voice communication technology in the online virtual world Second Life.

Voice technology for player-to-player communication is becoming a popular choice in multiplayer games, and has recently been introduced into the online virtual world Second Life. However, while voice may be well-suited to communication of fast-paced team games, it is less certain whether voice is suitable for non-gambling virtual worlds, where people meet mainly in order to communicate. This project will help us understand how voice is being used in Second Life, under what circumstances it is better or worse than communicating via typed text, and how it might be improved.

To study the use of voice in Second Life we want to ask you about your experiences using voice to communicate in Second Life. The interview should take about an hour. It can take place in person or by mail, depending on which is most convenient to you.

The project does not involve any risks. The only additional inconvenience for you is the time taken by the interview sessions and interviews. The information you provide will be treated as confidential and will only be used for research purposes connected with this project. Confidentiality of the information will be protected, subject to any legal limitations such as subpoenas or FOI. Access to the information will be restricted to the investigator only. As required by the University, data will be held in locked cabinets in the Department of Information Systems, and destroyed using confidential waste disposal techniques after five years following last publication from the research. No individual person or organization will be identifiable in the research report. However, due to the small number of participants in this research project, even though people will not be named, there is a possibility that people could be identified by research associates.

Your participation in the research project is voluntary, and you may decide to continue participating at any time. You may also withdraw any unprocessed data previously provided.

If you have any questions about this research project please contact Greg Wadley on 03 8344 1386 or greg.wadley@unimelb.edu.au or Martin Gibbs on 03 8344 1994 or martin.gibbs@unimelb.edu.au.

If you have any concerns regarding the conduct of the research please contact the Executive Officer, Human Research Ethics, The University of Melbourne 3018 ph: 8344 7507 fax 8344 0739.

---

Figure 36: PLS for study 4
The next document is the list of interview questions that was approved by the ethics committee.

THE UNIVERSITY OF MELBOURNE
DEPARTMENT OF INFORMATION SYSTEMS

<table>
<thead>
<tr>
<th>PROJECT TITLE</th>
<th>Using voice to communicate in Second Life</th>
</tr>
</thead>
<tbody>
<tr>
<td>INVESTIGATORS</td>
<td>Mr. Greg Wadley (<a href="mailto:greg.wadley@unimelb.edu.au">greg.wadley@unimelb.edu.au</a>)</td>
</tr>
<tr>
<td></td>
<td>Dr. Martin R. Gibbs (<a href="mailto:martin.gibbs@unimelb.edu.au">martin.gibbs@unimelb.edu.au</a>)</td>
</tr>
</tbody>
</table>

**Interview questions**

Our research involves individual and group interviews. These will be conducted using an open, semi-structured interview process. We will use the following questions to stimulate discussion; however discussion topics will also emerge from observation of play, and during the interview itself.

- How long have you been using Second Life?
- How much do you use it?
- What do you like to do there?
- How long have you been using voice-over-IP in Second Life?
- Did you like using voice-over-IP in Second Life?
- Was it better or worse than using text?
- How often did you use voice compared to text?
- In what particular situations in which either voice or text was preferred?
- How well was the voice communication implemented?
- Did your use of voice change during the study?
- How could voice be better implemented in online worlds?
- Were any aspects of the user interface for voice frustrating?
- Do you prefer a 'always-on' or a 'press-to-talk' interface for voice?
- How much have you used online worlds or MMOGs before?
- Have you used voice in games before? In what kinds of games? Was it useful?
- Have you used other voice-over-IP services such as
  - Is voice more useful in games such as MMOGs than in Second Life?
  - Is voice better for small or large groups of users?
  - Is voice better for person-to-person, vicinity, or group chat?
  - Is voice better for some people or 'personality types' than others?
  - Are there communication technologies other than voice that you would like to see introduced into online worlds?

*Figure 37: Interview questions for study 4*

During my first few months of using Second Life I kept a diary of my experiences regarding voice: here is an excerpt.
20 June

I just got voice working for the first time - I mean working as in I found some people I could talk with. I went to a few empty places an usual, then looked in Groups for “voice” and found one that listed some places with voice enabled. The first one I couldn’t find in Search/Places, but the second one was listed, so I tp there. There were about 20 avatars there in a semi-open landscape – this is more than I’m used to seeing. I pressed the push-to-talk key, tried to say something, and suddenly I could hear conversation. SL was extremely laggy, beyond laggy, closer to “frozen”, it reminded me of playing Quake in the 90s without a 3d card. My avatar movement was completely out of control – no movement at all for a while, then suddenly I traveled 50 metres, so I overshot the mark a lot moving around the area trying to find who was talking. Occasionally I drifted out of earshot – I’d guess the voice range is about 40 in-world metres. Three speakers were having a fairly intimate conversation (not as in sex but as if they knew each other well). Two females and one male. One English and two American. I could occasionally hear a young child crying or calling out in the background of one of the people speaking.

It took a good 20 minutes or so but I eventually figured out which of the avatars on the screen were speaking. In the meantime it was quite an odd experience – a conversation in my headphones that was quite unrelated to what I could see on screen, which was about 10 or 20 avatars standing around having typed-text conversation(e). I figured out roughly which part of the room the talkers were in because the voice got louder and softer as I moved around. At one point I was looking at three avatars who I thought must be the talkers, and one of them mentioned one of the other’s avatar-names, and then I knew I had them. I listened to their conversation for a few minutes – this felt slightly odd, I was eavesdropping in a way, but didn’t really feel like I had connected socially with them. One of them logged out, then another, then a couple of other avatars approached and joined in conversation with the remaining talker. At this point I cut myself as a voice by saying something. They replied and were fairly polite, but it seemed awkward, like they knew each other and not me. Without asking them, I had no way of knowing. Perhaps they realized I had been eavesdropping. Or maybe my voice sounds stupid. The experience obviously made me self-conscious about my skills at speaking with strangers, something I am not good at anyway. Using voice is a much stronger “social connection“, ie you feel a much stronger sense that the other people are people

Figure 38: Excerpt from my SL diary

I conducted a group interview on voice within a weekly Second Life discussion group. I have obscured Second Life usernames other than my own. (Usernames consisted of two names drawn from predefined lists: I have only obscured the second name.)
I conducted a number of individual interviews. Here is a sample of one of them. I anonymized this using text-replace in Word.

Figure 39: Excerpt of a group discussion from study 4
Figure 40: Excerpt of an interview from study 4

I regularly searched SL user forums and email lists for conversation about voice. Here is a sample forum conversation.

anon01: Social - My wife and I build our social network in SL, and then maintain it through the use of SL (and twitter - but that is OOS)
Grog Waydelich: They were referring to the way SL users often keep their SL nickname separate from RL name, don’t confuse them, keep SL presence separate
Grog Waydelich: Having met people in SL do you meet them later in RL or keep separate?
anon01: I (and her) also use SL for quite a deal of purposeful communication.
anon01: Where possible, I use SL as a social networking tool - similar to Blogs, and micro-blogs.
anon01: Yes, we have semi-regular SL/RL gatherings
anon01: I have now met >10 SL friends in RL, and that has improved my ability to understand the content of thier messages
Grog Waydelich: Broadly speaking, my research qestion is “how does using voice change the experience of online virtual worlds”
anon01: Voice _does_ help with understanding some of the communication overtones, but it also gets in the way as well.
Grog Waydelich: Ah -- could you expand on that last statement? --- a lot! :-) 
anon01: Ahhh, I think that it broadens, but reduced the depth of the communication
anon01: This is the difference between a movie and a book.
Grog Waydelich: That’s a nice analogy
anon01: When you have a book, the reader has to synthesize all the images, this process acts to bond those images and impressions deeper into the user/reader’s world-view.
anon01: When the images are provided from an external source, the user is simply a consumer, not a co-creator.
anon01: This to me, is the real strength, power, and danger of a visual+text environment
Grog Waydelich: Books and movies are quite different user-experiences .. does voice make that much of a difference in SL or other VW?
anon01: In a relationships/communication medium (such as SL is for a large number of participants), the addition of voice can be both powerful, and
anon01: restricting.
Dealing with voice grievers?

A public meeting place I maintain has developed a big voice following, however recently there have been reports of voice grieving. The island is no-build (which I hate, but autoreturn is no match against self-replicating screamers), so instead of rezzing garbage, griefers have resorted to finding out-of-the-way places to hide their avatars while they create disturbance in the voice channel.

If you can see a person who's speaking in voice you can tell that they're using voice, but they have been hiding in sneaky places, and in a crowded location the green dots don't always reveal where to look.

I'd like to hear from folks who run or visit public places where voice is (or was) common, and learn how grieving is handled. Right now I have a few 'deputies' who have banning privileges, but they still need to be able to track down the offenders, which isn't always easy.

Thank you.

Such a shame you have to deal with this.. I could have sworn we were on the adult side of SL.. As a rule.. I NEVER use voice.. there is a product avall that you can click and scan each avatar in the sim you are in and then tp to them.. That might be a solution you can try.

Finally I illustrate my data coding hierarchy with this screenshot from NVivo.
<table>
<thead>
<tr>
<th>Name</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abuse of channel</td>
<td></td>
</tr>
<tr>
<td>- etiquette of voice</td>
<td>3</td>
</tr>
<tr>
<td>- grieving, insults</td>
<td>7</td>
</tr>
<tr>
<td>Choice of voice or text</td>
<td></td>
</tr>
<tr>
<td>- few people using voice</td>
<td>5</td>
</tr>
<tr>
<td>- group choice, text as default option</td>
<td>1</td>
</tr>
<tr>
<td>- love or hate voice</td>
<td>8</td>
</tr>
<tr>
<td>Communication richness</td>
<td></td>
</tr>
<tr>
<td>- bandwidth richness</td>
<td>11</td>
</tr>
<tr>
<td>- speed of communication or explanation</td>
<td>2</td>
</tr>
<tr>
<td>Identity and social presence</td>
<td></td>
</tr>
<tr>
<td>- accents</td>
<td>1</td>
</tr>
<tr>
<td>- Immersion and Augmentation</td>
<td>12</td>
</tr>
<tr>
<td>- knowing who one’s interlocutor is</td>
<td>9</td>
</tr>
<tr>
<td>- personality, introversion, shyness</td>
<td>8</td>
</tr>
<tr>
<td>- role-play, gender, anonymity</td>
<td>12</td>
</tr>
<tr>
<td>- suspicion of voice-reducers</td>
<td>5</td>
</tr>
<tr>
<td>Leaks</td>
<td></td>
</tr>
<tr>
<td>- physical world sound leaks into virtual world</td>
<td>5</td>
</tr>
<tr>
<td>- virtual world sound leaks into physical world</td>
<td>12</td>
</tr>
<tr>
<td>Multiple simultaneous conversations</td>
<td></td>
</tr>
<tr>
<td>- large conversation groups, turn-taking</td>
<td>3</td>
</tr>
<tr>
<td>- multi-threaded conversation, multi-tasking, multiple modes</td>
<td>19</td>
</tr>
<tr>
<td>Technicalities</td>
<td></td>
</tr>
<tr>
<td>- disabilities, accessibility</td>
<td>1</td>
</tr>
<tr>
<td>- hardware and technical problems with voice</td>
<td>16</td>
</tr>
<tr>
<td>- problems with spatial audio</td>
<td>3</td>
</tr>
<tr>
<td>- store and search, transcription, recording</td>
<td>6</td>
</tr>
<tr>
<td>- typing skills, usability issues</td>
<td>6</td>
</tr>
<tr>
<td>What voice is used for</td>
<td></td>
</tr>
<tr>
<td>- special applications such as language teaching, rehearsing plays</td>
<td>14</td>
</tr>
<tr>
<td>- workplace meetings and telecommuting</td>
<td>4</td>
</tr>
</tbody>
</table>

*Figure 42: Data coding hierarchy from study 4*
# Appendix H: Virtual worlds cited

<table>
<thead>
<tr>
<th>System name</th>
<th>Release date</th>
<th>Publisher</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active Worlds</td>
<td>June 1995</td>
<td>ActiveWorlds, Inc.</td>
</tr>
<tr>
<td>Anarchy Online</td>
<td>June 2001</td>
<td>Funcom</td>
</tr>
<tr>
<td>Counterstrike</td>
<td>November 2000</td>
<td>Valve</td>
</tr>
<tr>
<td>Club Penguin</td>
<td>October 2005</td>
<td>Club Penguin Entertainment</td>
</tr>
<tr>
<td>Dungeons and Dragons Online</td>
<td>February 2006</td>
<td>Atari</td>
</tr>
<tr>
<td>EVE Online</td>
<td>May 2003</td>
<td>CCP</td>
</tr>
<tr>
<td>Everquest</td>
<td>March 1999</td>
<td>Sony</td>
</tr>
<tr>
<td>Habbo</td>
<td>January 2001</td>
<td>Sulake Corporation</td>
</tr>
<tr>
<td>LambdaMOO</td>
<td>1991</td>
<td>Xerox PARC</td>
</tr>
<tr>
<td>Second Life</td>
<td>June 2003</td>
<td>Linden Lab</td>
</tr>
<tr>
<td>There</td>
<td>October 2003</td>
<td>There, Inc.</td>
</tr>
<tr>
<td>Warhammer 40k (tabletop game)</td>
<td>1987</td>
<td>Games Workshop</td>
</tr>
<tr>
<td>World of Warcraft</td>
<td>November 2004</td>
<td>Blizzard Entertainment</td>
</tr>
</tbody>
</table>

[End of appendices.]
Author/s:
Wadley, Gregory Robert

Title:
Voice in virtual worlds

Date:
2011

Citation:

Persistent Link:
http://hdl.handle.net/11343/37224

File Description:
Voice in virtual worlds

Terms and Conditions:
Terms and Conditions: Copyright in works deposited in Minerva Access is retained by the copyright owner. The work may not be altered without permission from the copyright owner. Readers may only download, print and save electronic copies of whole works for their own personal non-commercial use. Any use that exceeds these limits requires permission from the copyright owner. Attribution is essential when quoting or paraphrasing from these works.