University Library



A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Wadley, G;Carter, M;Gibbs, M

Title:
Voice in Virtual Worlds: The Design, Use, and Influence of Voice Chat in Online Play

Date:
2014-12-22

Citation:
Wadley, G., Carter, M. & Gibbs, M. (2014). Voice in Virtual Worlds: The Design, Use, and Influence of Voice Chat in Online Play. Human–Computer Interaction, 30 (3-4), pp.336-365. https://doi.org/10.1080/07370024.2014.987346.

Persistent Link:
https://hdl.handle.net/11343/54786

Voice in virtual worlds: The design, use and influence of voice chat in online play

Running Head: Voice in virtual worlds

ABSTRACT

Communication is a critical aspect of any collaborative system. In online multiplayer games and virtual worlds it is especially complex. Users are present over long periods and require both synchronous and asynchronous communication, but may prefer to be pseudonymous or engage in identity-play while simultaneously managing virtual and physical use contexts.

Initially the only medium offered for player-to-player communication in virtual worlds was text, a medium well-suited to identity-play and asynchronous communication; less so to fast-paced coordination and sociability among friends. During the past decade vendors have introduced facilities for gamers to communicate by voice. Yet little research has been conducted to help us understand its influence on the experience of virtual space: where, when and for whom voice is beneficial, and how it might be configured.

To address this gap we conducted empirical research across a range of online gaming environments. We analyzed our observations in the light of theory from human-computer interaction (HCI), computer-supported cooperative work (CSCW) and computer-mediated communication (CMC). We conclude that voice radically transforms the experience of online gaming, making virtual spaces more intensely social but turning them into maelstroms of impression management, identity play, and ambiguity over what is being transmitted to whom.

CONTENTS

1. INTRODUCTION
    - 1.1. Communication in play
    - 1.2. Prior work
    - 1.3. Our research program
2. CASE STUDIES
    - 2.1. Study 1: Speaking with strangers in *Xbox Live*
    - 2.2. Study 2: Voice in massively multiplayer role-playing games
    - 2.3. Study 3: Propagating voice spatially in a team shooter
    - 2.4. Study 4: Using voice to collaborate around objects
    - 2.5. Study 5: Voice in a pseudonymous social world
    - 2.6. Study 6: Spatial voice and moral dilemmas in *DayZ*
3. DISCUSSION
    - 3.1. Voice transforms the player experience of online gaming
    - 3.2. Advantages and challenges of voice

3.3. Media richness and virtual worlds
3.4. Fictional social presence
3.5. Multi-tasking and multi-contexting
3.6. Simulated telecommunication in simulated space
3.7. Technical difficulties
3.8. Implications for design
3.9. Conclusion
4. BIBLIOGRAPHY

1. INTRODUCTION

1.1 Communication in play

Digital games are often social experiences. In homes, arcades, LAN parties and eSports stadia they provide a spectacle, a focus for conversation, and an opportunity for cooperation and competition. Collocated play affords players and their audience forms of communication such as speech and body language, enabling very sociable experiences in which people discuss plans, bark orders or laugh at each others' misdeeds (Aarsand and Aronsson 2009).

By contrast, the first generation of Internet-based virtual worlds (such as *Ultima Online*, *Anarchy Online* and the text-based MUDs from which graphical MMORPGs evolved) lacked the intense sociability of collocated play, in part because, while they afforded limited gesture and proxemics through the use of avatars, linguistic communication was restricted to "text chat". In these socially opaque environments a rich role-playing culture evolved (Raybourne 2001)[1] which incorporated a significant social dimension (Schiano 1999). However, as online gaming developed players also found that text did not afford the same levels of conviviality as collocated play, and was not an ideal way to communicate in the heat of battle (Wadley et al. 2003). Thus when gaming platforms and networks became sufficiently powerful, many players readily adopted "voice chat". Voice became available first via third-party products like *TeamSpeak*, *RogerWilco*, *Ventrillo* and *Mumble* and later via voice facilities included as part of games such as *Counter-Strike* and platforms such as *Xbox Live*.

Voice in games is now commonplace. Indeed, it is often considered mandatory by players and game designers alike. But how has it changed the experience of online gaming? What advantages and disadvantages does voice have compared with text? Does voice offer superior communication to text as suggested by some media theorists or are there circumstances when the leanness of text is preferable? What difficulties do people experience with using voice in online environments? How can studying the introduction

[1] Throughout this paper we will use the term 'role-playing' to refer to the creation and enactment of a fictional identity or character in a virtual world that typically involves 'acting out' the character using theatrical techniques. This should be distinguished from the other uses of the term to describe playing a functional role in a team, or the genres of role-playing games and massively multiplayer role playing games that use an avatar and place the player in the role of the game's protagonist: these may or may not involve role-playing as defined above.

of voice to virtual worlds contribute to our understanding of online communication more generally?

In this paper we reflect on over a decade of research that has comprehensively examined the implementation of voice communication in online worlds, and the impact this has had on gameplay and player experience. Our case studies include racing and first-person shooter (FPS) games in *Xbox Live*, role-play and raiding in *World of Warcraft*, spatial team play in *Wolfenstein: Enemy Territory*, collaboration in *Second Life,* and social negotiation in *DayZ*. Our work is situated at the intersection between HCI and game studies, and examines how technology configuration impacts the experience of simulated game spaces. It extends earlier work on collaboration in virtual environments such as that surveyed by Churchill et al. (2001), Schroeder (2002) and Schroeder and Axelsson (2006), and is concerned more with the *sociability* than the *usability* of voice in virtual environments (Preece and Maloney-Krichmar 2003). Our work complicates the view that voice is richer and thus necessarily better than text for communication in virtual worlds. It also exemplifies the utility of contemporary HCI's ethnographic stance (Dourish 2006; Harrison et al. 2007) for creating deep understanding of the impact of technological change upon user experience.

1.2 Prior work

Online multiplayer videogames and especially massively multiplayer role-playing games (MMORPGs) and virtual worlds (VWs) offer excellent examples of online collaboration among virtual teams. Many require teamwork that is intensive, detailed, sometimes asynchronous and sometimes fast-paced in order to achieve game goals (e.g. Golub 2010). Some of the situations encountered in games are analogous to those faced by, for example, soldiers and emergency workers, who have (not unproblematically) adopted two-way voice radio for its support for the coordination of small teams who are moving through space (cf. Toups et al. 2011). Yet the range of activities carried out requires that asynchronous coordination is sometimes employed. These complex engagements have meant that virtual worlds have often been studied as CSCW systems (Ducheneaut and Moore 2004, Moore et al 2007, Nardi and Harris 2006). The sheer scale of some virtual worlds, in terms both of number of players and size of virtual space, as well as the rich player culture they emerge, have also made them rich sites for ethnography (Nardi, 2010; Boellstorff, 2008).

However virtual play spaces are at the same time unique among CSCW systems. Players carry out their tasks not in physical space but a simulated one (Aarseth 2008), and are represented (or "embodied") as avatars (Bailenson and Blascovich 2004). The spaces are usually fictitious, often inspired by science fiction or fantasy themes, and the avatars resemble imaginary characters that match the game's theme or back-story. This suggests that users are role-playing, a concept which is reflected in the genre name "massively multiplayer online role-playing game". Definitions of role-play have varied from playing a functional role in a team to enacting a fictional character (Yee 2006, Neustaedter and Fedorovskaya 2009). We adopt the latter approach and use the term 'role play' to denote the acting out of a fictitious identity. Players may be able to customize their avatar's appearance or choose demographic properties such as gender and race; this supports role-

play and a degree of presentation-management (Ducheneaut at al. 2009), although more recent work in HCI has begun to critique the importance of avatar appearance in the experience of virtual worlds (Carter et al. 2012a).

Early social presence (Short et al. 1976) and media richness (Daft et al. 1987) theories held that "rich" (audio-visual) media project cues to the identity and state of the person communicating, such as gender, age, nationality, education level and mood, while "lean" (textual) media project little such information. While these theories have been considerably refined, it is clear that to some extent that users select communication media with these properties in mind. For example, people choose text over voice when eavesdropping is a concern, or when they desire more control over self-presentation (Walther, 1996; Carlson et al., 2004).

Therefore it is plausible that the communication media used by gamers significantly influences their activities in virtual worlds, and that gamers will choose amongst available communication media depending on the experience they want to have. Yet until our work began, relatively little research had addressed mediated communication within virtual worlds. Sallnas (2002) compared decision-making by VW users equipped with text, voice, or a video link. Nilsson et al. (2002) supplemented a text-only virtual world with a shared-audio system to study workplace meetings held in the virtual world. Halloran et al. (2004) studied the use of voice by players of first-person shooter (FPS) games, finding that voice could increase sociability and free up of hands for movement control, but that benefits were situation-dependent. Williams et al. (2007) found that MMORPG teams who used both voice and text to communicate over a period of months liked and trusted each other more and were happier than those who communicated only by text. Other than this work (and the work reported in this paper) there has been little systematic investigation of the influence that the voice medium has had on the experience of virtual world users.

A thorough and comprehensive account has not previously been made of what advantages voice brings, or conversely which conditions, if any, render voice unsuitable, nor how this might be addressed by designers. Without understanding the experience of voice users, developers run the risk of failing to successfully implement these large, expensive systems. This is evidenced by the limited success enjoyed by some implementations; for example, the voice system added to *World of Warcraft* has been rejected in favor of third-party voice systems (Street 2011), and the introduction of voice to *Second Life* was also met with controversy - some players not only refused to use voice but even threatened to quit altogether if it was introduced (Boellstorff 2008). Virtual world communication mechanisms need to be usable by a large, heterogeneous player base, and the design space for voice is potentially broad, yet has only been superficially investigated. Therefore an understanding of how different configurations are received by different game players conducting activities in different contexts will be of value to both developers in the games industry and HCI research more broadly.

1.3 Our research program

Given this gap we have chosen as our research question: *How does voice influence the user experience of virtual worlds?* To understand the use and influence of voice in online games we have conducted empirical research across a range of gaming contexts and virtual spaces (Wadley and Gibbs 2010), from console (Gibbs et al. 2004, Wadley et al. 2005) and PC based team shooters (Gibbs et al. 2006) through several MMORPGs (Wadley et al. 2007), *Second Life* (Wadley and Ducheneaut 2009, Wadley et al. 2009), and the recent massively multiplayer online first person shooter (MMOFPS) *DayZ* (Carter et al. 2012b, 2013).

As Schroeder (2002) argued, a fundamental problem in studying virtual worlds is choosing a site for observation, seeing as the action takes place in both physical and virtual contexts. This was particularly relevant for our project, because the use and suitability of communication media partly depends on context. Whereas some VW researchers have conducted "virtual ethnography" (e.g. Moore et al. 2007, Boellstorf 2008) we elected to gather data in both physical and virtual contexts.

To develop a deep understanding of user experience we used methods drawn from ethnography such as individual and group interviews, diaries, observation, interaction analysis of screen recordings, questionnaires and analyses of player discussion forums, and analyzed the rich, qualitative data thus gained by using thematic analysis. In this we followed a tradition of using ethnographic methods to understand use of virtual worlds (e.g. Taylor 2006, Nardi and Harris 2006, Ducheneaut et al. 2006, Boellstorff 2008). We have interviewed people with diverse approaches to game playing, and have observed, participated and recorded many play sessions across genres and platforms. Detailed accounts of data collection methods used can be found in (Gibbs et al. 2004, Wadley 2011; Carter et al. 2012, 2013). All studies were conducted according to national guidelines for ethical research involving humans, and all were approved and monitored by the University of Melbourne Human Research Ethics Committee.

We recruited participants from diverse populations and with a range of experience with the technologies in question. 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. The technology used in all cases was commercially available, except for one voice system which was under development during the study and has since been commercially released. Figure 1 illustrates the timing of our studies with respect to relevant innovations in gaming:

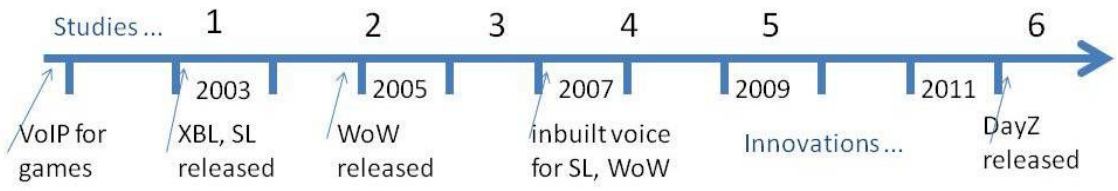


Figure 1: Timeline of studies

Grounded theory (Glaser and Strauss, 1967) informed our overall approach, in particular selection of cases and interpretation of data. Grounded theory favors the discovery of new theory over verification of existing theory, and is particularly well suited to exploratory studies of new phenomena (Stebbins, 2001). The principle of theoretical sampling, which seeks to maximize the opportunities for comparison within a limited sample, guided our choice of study platforms. We chose *Xbox Live* for our initial research as it was the first gaming platform to include (indeed, mandate) voice chat as standard. Because this study highlighted speaking with strangers and large groups as important challenges for voice, we then chose massively multiplayer games as sites for research. We studied an early prototype of spatial voice as a contrast to the "radio" style voice configuration of our first two studies. We then examined voice chat during object-focused collaborative tasks at the micro scale of virtual space. Because the need for rapid team coordination, and difficulties with role-play, had emerged as important themes in our earlier studies, we examined the introduction of voice into *Second Life*, a virtual world in which role-play but not combat is significant. Finally we studied a unique implementation of voice which enables combat with unknown players. Thus we have reflected on the uptake and evolution (or stagnation) of voice over most of the period that it has been used by gamers (Wadley 2011, Wadley and Gibbs 2010).

Further details regarding the methods of each individual study are noted in their corresponding sections. We coded data using the *NVivo* software package: the themes discovered are presented in the next section.

2. CASE STUDIES

Our research was structured as a series of case studies in which we examined how voice was used and how it impacted use in a variety of game genres. We began in 2003, when voice had limited uptake, by observing use of the *Xbox Live* network, which introduced voice-enabled online console gaming. This section describes each case study and shows how voice for gaming has evolved over a decade.

2.1 Study 1: Speaking with strangers in *Xbox Live*

Introduction

The *Xbox Live* (XBL) network, released in 2003, was the first successful system for connecting hitherto standalone game consoles to the Internet to enable multiplayer videogaming over a distance. Among XBL's novel features was a 'gamertag' system for the central management of consistent online identity across games. In conjunction it offered a voice communication system and included a microphone and headset as part of the subscription package. *Xbox Live* (initially and for several years after this) included no means of typing messages, and developers were compelled to support voice in all games. At the time of launch, text was still the dominant medium for communication in online computer games, with third-party voice products having just begun to appear. This novel combination of features inspired us to conduct a study of *Xbox Live*.

Method

Our research was conducted with six groups of three volunteers. Each group participated in one, two-hour research session. At the start of each session, participants completed a background questionnaire. They were then asked to play multiplayer games with other players over the Internet using *Xbox Live*. Participants were aged between 18 and 27 years and both genders were represented. Participants reported that, on average, they played PC-based computer games more often than console games (6 hours per week versus 2 hours). Two participants played for over 20 hours per week. Only three participants played console games more often than other computer games. Participants liked several different types of game, with the most popular being first-person shooters, action games, sports, strategy games and role-playing games. All participants owned or regularly used a computer and had Internet access.

Two games were used in this study: *MotoGP* and *Unreal Championship*. *MotoGP* is a motorbike racing game that was packaged with every *Xbox Live* starter kit. *MotoGP* incorporates a 'lobby' area, outside the main racing arena, where players waiting for their race to start can converse in a party-line fashion. During a race, a proximity-based algorithm is used, allowing players to talk to each other if they are close together on the track. *Unreal Championship* is a first-person shooter allowing either 'death match' (all-against-all) or team-based modes of play. In a death match, each player can hear all other players speak, while in the team games, all players on a given team can speak to and hear each other independent of their location in the game world.

We observed the study participants interact with other players via the network and other participants in the room. After each group of participants played games for an hour, a researcher led them in a focus group discussion. Open-ended questions sought the participants' opinions about their first encounter with *Xbox Live*. Questions focused on topics such as the voice headset, and playing online. The sessions were recorded using a single video camera directed at the participants (the gamers with whom our participants interacted online were not recorded). This data was analyzed to identify common themes across the groups through a process of open-coding, then closed-coding to refine and confirm prevalent themes.

Groups were used in the observation sessions to encourage participants to express their views and describe their experience. This approach was based on the constructive interaction technique developed by O'Malley et al. (1984) for usability testing. The focus group discussions that followed each game playing session provided participants with an opportunity to discuss and reflect upon their initial encounter with *Xbox Live*. These discussions were used to provide insights into participants' opinions, feelings and attitudes about their experience playing with *Xbox Live*.

Findings

Participants valued the social interactions entailed in playing with other people. According to one participant, "*It's the interaction with other people and it's the buzz playing with other people and if you're really competitive and you beat the other person you feel good about it."* Participants also praised the convenience and usability of voice communication for gaming. They felt that voice was a step forward from traditional text-

based messaging: "*The headset is good in tactical swat team games. That's really good because that's really fun. Like at LAN parties and stuff, you shout across computers and stuff like that."* In particular, many of our participants thought that voice would be beneficial in tactical team-based games, by better supporting the coordination of team members. An important advantage cited was the freeing up of hands from the task of typing messages "*I quite like the headset, the fact that you can talk. Especially in Unreal where, well the fact that you're supposed to work as a team and you can sort of do that quite well. I think it's better than having to type or anything like that."*

XBL was the first voice-only game network and both participants and researchers had expected a highly sociable experience. However this was often not the case. Players were placed into game sessions with people they did not know, and were often for this reason reluctant to use voice. The voice channel was marked by verbal abuse and the broadcast of noise such as recorded music. Four types of "noise" were present on *Xbox Live*. First, participants could hear conversations that they were not a part of. These conversations were just as prominent in the voice channel as were conversations directed at the participants. Second, participants overheard offline conversations between groups of players collocated around a distant console on the network. By leaving the microphone active, part of that conversation would be broadcast over the network. As one participant commented, *"It just sounds like they are talking to themselves. I mean, we [gestures to other participants in the room] are talking to ourselves."* Third, participants encountered sound that was unintentionally transmitted by other players, such as television, people making motorbike sounds, and background music. Fourth, sometimes these sounds appeared to have been sent intentionally: a form of griefing. Examples included loud incoherent speech, relentless trash-talking, noise-making devices such as mobile phones, and music at such a volume that the sender's headset microphone had apparently been held close to a hi-fi speaker.

Participants also had problems identifying who was talking on the voice channel, finding themselves unable to link voices they heard via the headset to either gamertags or avatars in the game: "*It's hard to identify who you are talking to. I mean you can't get the sense of community if you are connected to all these people, you can't really see them or you don't know who you are talking to."* This was particularly pronounced in Unreal Championship, where the voice channel was configured in a 'radio' mode such that everyone on the same team could hear each other equally, regardless of their position in the game world. This effectively 'disembodied' the voice.

On first exposure, participants experienced the voice channel as chaotic and out of control. They desired the ability to direct their messages to specific people: *"I didn't know who I was talking to. There wasn't functionality to select who I was talking to."* They were uncertain whether their utterances were heard by the intended recipients. One was heard to say: *'I'm hanging around the flag. Where are you? Can you hear me? Hello?'* During other sessions participants repeatedly asked: *"Is anyone there?"* or *"Hello?"* to prompt a response; this interrupted the social experience and flow of gameplay.

Summary

Our first study of voice in gaming was an exploratory first look at a novel new gaming platform and in particular at the voice communication with strangers that it enabled and to an extent mandated. Until the appearance of *Xbox Live*, most online communication among gamers had been via typed text. Voice, when used at all, was restricted to small groups of friends who knew each other and often played together. XBL's bringing together of ad-hoc play groups allowed us to examine the experience of being exposed to voice with strangers. While the advantages of voice were clear, especially during fast-paced action involving the coordination of teams, this global voice-enabled online network was not the usability and sociability nirvana one might have expected. The lack of sufficient contextual clues to identify who a speaker was, the bleeding of non-game and irrelevant noise and conversation into the game world and the reluctance of some players to participate in voice conversation stymied its successful implementation.

This study was limited in that use was not examined over an extended period, and we did not explore in depth issues such as the interaction of gender with voice chat (for recent work exploring the heavily gendered experience of *Xbox Live*, see Xu et al., 2011; Kuznekoff, 2012; Gray, 2013). However it provided the insight that implementing voice in games was not necessarily a straightforward improvement. This inspired us to undertake more studies, in different types of games, with different types of players, and with differently-configured voice channels. In particular, it appeared that not knowing to whom one was speaking was a major barrier to the acceptability and success of voice, while coordination of teams during fast-paced action was a major draw. Thus for our next study we examined voice in the MMORPG genre, in which different kinds of game play demanding coordination occur over longer periods, involving players both known and unknown to each other.

2.2 Study 2: Voice in massively multiplayer online role-playing games

MMORPGs are complex online worlds. They are persistent (individual game worlds may last for years) and typically host thousands or even millions of players simultaneously, so that most do not know each other except through the game, and collaborators are usually not physically collocated (Yee, 2006). Players often form long-term associations ('guilds') which are typically larger than FPS teams and often hierarchical in structure (Williams et al., 2006). They collaborate in team-based game scenarios, the most notable of which is the 'raid' – a joint attack on an AI monster which offers significant rewards and cannot be performed by a lone player or uncoordinated teams (Golub 2010, Chen 2012). Players may also form temporary associations called 'pickup groups' which are usually smaller and may involve strangers not drawn from their guild. Players can also engage in player-vs-player (PvP) battles, and spend time exploring the virtual world, socializing with guild mates and strangers passing by, and collecting and trading virtual possessions (Ducheneaut et al. 2006). While many players are highly focused on achievement, for many the game is also a social experience (Steinkuehler and Williams 2006; Ducheneaut et al. 2007) which makes the ability to conduct convivial communication particularly important.

Prominent game designer Richard Bartle sparked debate in 2003 by predicting that the burgeoning popularity of voice chat among FPS players might spread to MMORPGs 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 2003). His prediction of wide adoption proved to be correct. By 2005, *World of Warcraft* (*WoW*) had become the most popular MMORPG in the US, partly by attracting a young audience that also played FPS games or the RTS game Warcraft from which *WoW* emerged. Many of these brought existing practices including the use of third-party voice products, which became mandatory for membership in most end-game raiding guilds (Golub 2010).

Following the enthusiastic adoption of third party voice services by *WoW* players, *Dungeons and Dragons Online* (*DDO*) in 2005 became the first MMORPG to be released with voice chat integrated into the game software. Several other voice-enabled MMORPGs followed, such as *Lord of the Rings Online* (*LOTRO*), released in 2005/6, and *EVE Online*, released in 2003 and voice-enabled in 2007. A built-in voice channel was added to *WoW* in 2007, competing with third-party products. Implementing voice within the game allowed channel management to be integrated with team management, and allowed the game client to visually indicate use of the channel, ostensibly to solve the kind of problems identified in our first study regarding disembodiment of voice.

The *DDO* voice channel was configured such that when individual players joined a pickup group they could speak with team-mates in a "two-way radio" format. Players could not speak with non-team-mates nor with ex-team-mates after the team had disbanded. *DDO* did not distinguish between pickup groups and guilds, though it supported long-term associations with a friends-list feature.

Method

We arranged for three groups of five participants to use voice chat in an MMORPG over a period of two months. All participants were experienced MMORPG players, and many had already used voice in online FPSs. Two of the participant groups played *Dungeons and Dragons Online*, the first MMORPG with built-in voice chat. The members of one of these *DDO* groups knew each other prior to the study and had frequently played other online games together as a group. The other *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. None of the *DDO* participants had played this game before. A third group were already regular users of third-party voice products in *World of Warcraft*, *Everquest* and other MMORPGs: they continued to do this during the study and contributed data.

Participants played in their own homes, under their normal playing conditions. They kept diaries in which they recorded feelings and opinions about their use of 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 groups. Interviews and focus groups were open and semi-structured and lasted one to two hours. We sought to understand the players' use of voice and to collect their criticisms of existing

implementations and suggestions for future ones. We used canned questions to seed the discussions but allowed conversations to flow according to players' interests. For example we asked whether participants preferred voice or text, whether either medium was better suited to particular types of gameplay, whether they had encountered particular episodes in which voice was especially useful or problematic, and whether there were aspects of the voice interface they would like changed. We devised six fictitious gameplay scenarios for the focus groups and asked the participants to consider how they could use existing or imagined voice systems to deal with each scenario. We analyzed the participants' responses and organized them into themes. These are presented in the next section.

Findings

Nearly all of the participants approved of communicating by voice – under the right conditions, which they elucidated. After technical problems were resolved, participants found speaking to be easier, more natural, and more relaxing than typing, and found it added a new social dimension 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."* 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."*

Participants emphasized that it was interaction with other people that drew them to online virtual worlds. For some, voice enhanced social interaction: "*It's not just functional, it's social."* Those playing with friends liked the greater social presence conveyed by voice, saying it made the game "*feel like a living breathing party"*, with one likening the experience to a LAN party (i.e. collocated gaming). Some felt that the use of voice helped bind guilds together by making them feel they knew each other. In fact, players who didn't use voice became "*invisible*" to the group. *"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."* Voice enhanced "*off topic*" socializing unrelated do the game, reminiscent of the casual, often asynchronous chat employed by users of push-to-talk cellular radio (Woodruff and Aoki 2003): "*We use it for just chatting about stuff we've done in the daytime, and having a joke, that sort of stuff."* However not all participants appreciated off-topic talk: a chatty player could easily dominate the channel and annoy team-mates with what one participant described as "*verbal diarrhea*".

All agreed that voice was superior to text during raids. Because speaking freed a player's hands from typing, players could communicate with team-mates while carrying out other game actions such as moving and fighting. Some reported incidents in which this made the difference between success and failure. Coordination was especially important among large teams. Voice was a better way for raid leaders to issue directions to subordinates, and for wounded players 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.

Voice's immediacy and the extra presence it conveyed increased the emotional impact of messages. This could be a force for good in team management. On the other hand, the immediacy and emotional impact of voice could lead to more transmission of criticism and negative emotion. Voice made criticism feel more "*real*" to the receiver. 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. 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 is harder if not impossible in voice.

Several participants noted that they could discern aspects of the identity of fellow players by listening to their voice: "*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"*. Players adopting voice sometimes realized that they had been grouping with children. Sometimes the effect was merely comical, as when the voices of young children mismatched the appearance of their hulking warrior avatars. (One participant, a school teacher, pointed out that revealing the youth of players online was potentially dangerous.)

Participants debated whether role-playing could be conducted if voice made their personal characteristics more discernible. There seemed to be a widespread understanding that role-playing required the relative leanness of a text medium. One 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. With voice, it was felt that role-playing extended beyond what was said, to how it was said, which required some form of acting skill, the use of accents, vocal mimicry and the like. However role-playing 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-playing or even witnessing it, and seemed to perceive a stigma attached to role-playing. Some discussed why role-playing in voice was possible "face-to-face" (such as in tabletop games like *Dungeons and Dragons*) and yet problematic online. Again the concern seemed to be that fellow players on the Internet were strangers and possibly eavesdropping. Although most participants appreciated the advantages of voice, some felt that an MMORPG lost some of its atmosphere when players could hear each other speak: *"You lose the magic to some extent."* For some players, part of the appeal of virtual worlds is the *limited* social presence of other people.

In discussing use of voice, an issue frequently raised by participants was the choice of an avatar of the opposite gender. Confirming Yee (2006) and Blinka (2008), this usually meant males playing females avatars. Men who played female avatars did not perceive this practice as problematic, and most discussed it in a matter-of-fact manner. The reasons they gave for the 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* players' use of opposite gender avatars as weird or uncomfortable, and this may reflect the unease around cross-gendering suggested by MacCallum-Stewart. Some simply found that gender-swapping while using voice chat made the job of matching the voices that one heard with the avatars that one saw more difficult. One female participant said she used a female avatar specifically so that her voice and appearance would match. Participants who engaged in gender-swapping spoke of problems that arose when they adopted voice and exposed their offline gender to others. This discomfort seemed to depend on whether the player knew the people they were engaging with, which suggested some degree of anxiety about being "found out".

Unease arose in situations where voice-using players could be overheard by people they didn't know. Some participants found speaking with strangers to be uncomfortable. A viewpoint frequently offered was that voice was better for friends who collaborated over long periods than for pickup groups meeting for the first time, as reported by participants in study one. This depended on personality, with players self-identifying as shy particularly reluctant to use voice when teaming with strangers.

Participants' concern with 'who could hear what' went beyond the transmission of voice to include sounds that originated within their physical environment. As seen in the XBL study, participants described numerous incidents in which sounds such as breathing, eating, household noise, speech from family members, TV and music were accidentally transmitted into the voice channel. A participant described a slightly embarrassing situation in which he paused playing in order to go and comfort his small child, placed his headset beside a baby-monitor speaker and broadcast the scenario into the game. One participant reported overhearing younger players being told by their mothers to get off the computer. Conversely, participants noted that speech and other sounds intended to be heard within the game could be overheard by people who were collocated in a player's physical context; a problem that cannot arise with text (unless someone were to read the player's screen). This issue was typically resolved by wearing headphones (which has become a common practice amongst gamers). Two others reported that they played and spoke differently depending on whether family members, especially children were at home and able to hear them.

While voice chat provided raiding parties with coordination advantages similar to those provided by two-way radio, it also imposed some of the problems faced by radio users (cf. Camp et al. 2000, Toups et al. 2011). Participants complained that voice channels were prone to congestion far more than were text channels. The larger the group playing, the less useful was voice, because people became more likely to talk over each other, unless some form of 'discipline' over the use of the voice channel was accepted by all players (see also Bowers et al. 1996, Nilsson et al. 2002). A congested text channel, by contrast, was still usable because of the way text utterances were parceled into discrete sentences and persisted on-screen (see also Brown and Bell 2004, Landgren, 2006). A participant reported that his group didn't try to use voice in groups larger than six, and felt that waiting for his turn to speak detracted from immersion in game action. Another reported that raid leaders who controlled the team's Ventrilo channel were prepared to mute team members who cluttered it.

Two-way radio can also suffer from channel congestion, and to solve this problem users have devised radio discipline and a stylized vocabulary and grammar similar to that used by specialist radio users (cf. Woodruff and Aoki 2003). Our participants discussed whether formal protocols were or should be utilized in MMORPGs. 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 desired and had experienced the disciplined use of voice by raiding parties, where strict, formal voice procedures was necessary for overcoming difficult obstacles (for a similar conflict in the MMOG *EVE Online*, see Milik, 2014). Some guilds reduced clutter by employing "officer channels" distinct from the main raid channel in a manner reminiscent of the system designed by Camp et al. (2000). Some participants felt a 'public' voice channel would be feasible for areas like taverns in which large numbers of players congregated to form groups, trade and so on, while others felt that channel congestion would prevent voice from ever being used for public chat.

MMORPGs typically offer several text channels, each of which follows different propagation rules: for example, 'guild chat' is broadcast to the entire guild, 'raid chat' to the group currently engaged in combat, 'vicinity chat' to players whose avatars are nearby in virtual space, while one-to-one chat is for private conversations between two individuals. Most felt it would be impossible to monitor and contribute to several voice channels simultaneously. But some participants reported playing with groups in which some members used text and the rest used voice: these tended to split into two sub-groups depending on channel so that it became impossible to coordinate the whole group. Voice-using groups often continued to use text as a backup for important messages, or for when voice wasn't working for some team members. Text forums on dedicated websites and email were the preferred media for asynchronous communication between game sessions.

Some participants reported problems identifying the speaker of an utterance. This problem does not occur in a text channel, because printed utterances are prefixed with the sender's username. The voice system integrated with DDO highlighted players' names when they spoke, however third-party voice clients could not do this.

Summary

MMORPG players perform identities and occupy locations in both virtual and physical spaces (Nitsche, 2008). Reality and fiction collide in conversation, which can expose players and their surroundings. Thus voice breaks down the barrier between physical and virtual worlds. The suitability of different media for communicating in MMORPGs depends on a dynamic compromise between the need for people to understand their collaborators and their desire for privacy and a fluid identity.

Voice chat made coordination of group actions such as raiding more efficient, and made online play more easy, fun and social. However the immediacy of voice made flaming more intense, potentially intensifying social friction. Gamers also needed to consider who could overhear them in their *offline* setting.

The most significant impact of voice upon MMORPGs though involved its conveyance of a speaker's identity into the online setting. Some of the anonymity and role-play that had characterized earlier persistent game worlds was lost with the introduction of voice. People could readily tell the gender, age and nationality of those they encountered online. Some players, notably those who used opposite-gender avatars or who played with those who did, found this breakdown of anonymity uncomfortable.

Virtual worlds are distinguished from other collaborative systems by their simulation of a 3-dimensional space (Aarseth, 2008). At the point we began our third study, voice systems for games had essentially ignored this spatiality, offering point-to-point or radio-style chat among players. But new voice systems were appearing which took spatial position into account, and it was to these that we next turned our attention.

2.3 Study 3: Propagating voice spatially in a team shooter

Some of the participants in studies 1 and 2 discussed the possibility of a voice channel that took avatar location into account, analogous to the proximity text channels that were offered in most MMORPGs. In such a channel, when someone spoke, their message would only be received by people whose avatars were near the speaker's. Other participants felt that voice should be directional in order to give cues to match voices to avatars. They felt these properties could solve channel clutter, prevent the disembodiment of voice, and make it possible for voice to be used by large groups rather than restricted to a predefined team.

At the time we began study 3, all existing voice systems in virtual worlds conformed to a 'radio' metaphor; communication was organized into channels so that all players logged into a given channel received the same utterances, unaffected by their actions or locations in the virtual world, as with a radio. Some researchers (e.g. Terrano, 2003) had experimented with a spatial mode that mimicked the transmission of sound in air. According to this metaphor, the volume at which an utterance was heard decreased with the distance between sending and receiving avatars. We took the opportunity to conduct a trial of one such system which had been created by Boustead et al. (2005).

SpatialVoice (*SV*) allowed game players to hear anyone close enough in virtual space, whether friend or foe. It was stereo-directional, in that utterances appeared to come from the speaker's virtual location. This required that the *SV* system have real-time awareness of avatar locations and be integrated somewhat with game software. *SV* also allowed players to switch temporarily to a 'radio' broadcast mode. At the time of this study, *SV* had been integrated with some team-based FPS games. We studied the use of *SV* by people playing *Wolfenstein: Enemy Territory (ET)*. This was a World War II themed, networked FPS in which two teams competed during individual game sessions that typically lasted 30 to 60 minutes. Upon login the players formed into 'Allies' and 'Axis' teams, one of which 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 perform different roles such as infantryman, engineer and medic. Success depended on coordinating these roles effectively, so good communication was critical.

Method

Nine employees at a large commercial institution (not a university or game vendor) were recruited to trial *SV*. All were males aged between 20 and 40, who already played a regular weekly lunchtime game of *ET*. Initially seven participants were involved in the study but two more joined after the first couple of weeks. They were all experienced players of *ET*. Most of them also had experience with using voice clients such as *Roger-Wilco*, *Teamspeak* or *Ventrillo* as part of their previous gaming experience, which enabled them to draw comparisons between *SV* and these other systems.

Participants were asked to continue playing their regular *ET* session, but to use *SV* to communicate while playing during the study. Two participants were 'shadowed' during each trial session; that is they were observed and videotaped during the session to collect information on how they set up *SV*, any difficulties experienced, how *SV* influenced their game play and their interactions with other players, and so forth. Each player was shadowed at least once. Each participant was also asked to keep a diary in which they were encouraged to record their experiences of using *SV*, whenever they had time to make an entry, e.g. between game-play sessions, after a session, at home, and so forth.

A pre-trial interview was held with all participants approximately one week before the trial commenced. Interview questions sought participants' demographic data, their history with game play and their experience with online playing and other voice technologies.

Post-trial, the participants handed back their diaries for analysis, and participated in an informal 'close off' session to discuss impacts on game play. The study lasted 4 weeks spanning 5 play sessions; however after the trial, participants chose to continue using *SV*. After several more months of use they participated in a more formal focus group.

Findings

These players, who worked together and knew each other well, overwhelmingly favored voice over text. Voice for them made their lunchtime game played in separate offices feel collocated and intensely social: *"Five seconds later you see your plan totally disintegrate. And then we're screaming at each other: 'Where are you going?' [laughter]".* They felt voice had become an integral part of their gaming experience: *""For the first time in a few months I tried to play ET online on the weekend, and I really missed voice. It was extremely quiet."*. They 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."* They felt team cohesion and planning had been enhanced: *"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."*

Early in the trial participants regarded the spatial mode of voice propagation as a restriction over the more familiar radio mode: they saw no advantage in only broadcasting to a subset of the team. However over time participants discovered advantages of the spatial mode and used the radio mode less often. Some liked the extra realism of their voice imitating sound in air. More practically, spatial voice helped to filter out irrelevant chat and reduce the traffic they received. *"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."* Likewise, warnings such as "two enemies round the corner" were received by people to whom they were most useful. Utterances like "go left" or "get the truck" are indexical: awareness of proximity allowed players to make sense of them, and voice was consequently less disembodied. Spatial propagation allowed small, non-overlapping conversations between groups of two or three people to fluidly form and disband easily as ad-hoc teams formed in virtual space. Channel congestion was reduced and multiple conversations enabled without the need to manually configure 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.

Some of the participants were MMORPG players, and speculated about the potential for spatial voice in large persistent, virtual worlds. They that felt *SV* could scale with group size and might be useful when raiding. They felt it might also work at "inns" and other public meeting places where many people were closely collocated. (However we would note that situations frequently arise in MMORPG play where communication is required at a distance.)

An interesting aspect of using spatially propagated voice was the ability to hear the voices of members of the opposing team. The ability to listen to and converse with the opposition inspired new game tactics. Some enjoyed taunting members of the opposing team, and some who were dying in battle used voice to distract the opponent who had just beaten them and report their positions to comrades. The ability to overhear enemy conversation added new sociability to gameplay but a new danger as well: the channel now had to be used with care as it opened the possibility for espionage and counter-espionage. Some tried to actively mislead the opposition team, such as by crying out in different voices while defending a building, to create the impression of greater numbers.

Summary

While spatial voice seemed at first to be a constraint over the more common radio mode, the constraint made some communication more intelligible. Players received not only the spoken message, but a sense of how close and where the speaker was, 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.

Voice systems that emulate radio are disconnected from the spatiality of virtual environments and by and large from the events occurring in them. Integrating voice with virtual space increased immersion and made achieving and maintaining communication between team members a significant part of team strategy, mimicking the problems of

physical space. However there was debate over whether such realism was always desirable. Some players perceived a trade-off between realism and utility. This is reminiscent of Golub's (2010) argument against the necessity for visual realism in massively multiplayer online games. Golub found that high-performing MMORPG players were willing to give up the detailed graphical view of virtual space offered by game software in favour of situational awareness add-ins that visualized player and enemy statistics in real time, if it helped them win battles. Our study indicated that adding realism to a voice channel (in the form of spatial propagation of sound) would only be accepted if it helped or at least did not hinder players' efforts to win,

Having examined one of the first voice systems to take into account the spatiality of virtual worlds, we then turned our attention to a type of collaboration that exploits spatiality at a different scale - manipulation and conversation around 3-dimensional objects.

2.4 Study 4: Using voice to collaborate around objects

Studies 1 to 3 were concerned with players negotiating the 'macro' spatiality of the distribution of avatars in virtual space. Here the fundamental problem is that avatars are dispersed so that it is difficult for players to keep track of each other visually to coordinate joint activity. 'Radio' configured voice mimics real-world solutions to this problem. But sometimes communication also needs to occur within the 'micro' spatiality of collaboration around objects. In the physical world, people gather around artifacts 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 where their foci of attention are: the objects and locations to which they are referring in speech.

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 physical space, collocated collaborators can acquire information about each other's activities and the state of their field of work with relative ease (Schmidt, 2002), using hands, arms, face, eyes and body to deliver non-linguistic signals including iconic, indexical, and demonstrative gestures (Clark, 1996). One might expect that virtual environments should support a similar kind of deictic (indexical) reference, since players are embodied as avatars which can move and gesture. However research in the 1990s showed that avatar-mediated reference to objects in virtual spaces often failed due to problems with understanding spatial reference (Bowers et al. 1996, Hindmarsh et al, 1998).

We were interested in whether these problems persisted in modern virtual worlds, because there are several reasons why the situation might have improved, such as more realistic graphics, improved avatars, and greater familiarity among the population with 3D software. We decided to study this phenomenon in *Second Life* (*SL*), a popular virtual world which allows players to collaboratively create and modify objects. *SL* has a large player base and a shared culture relating to building objects. It offers an inbuilt voice channel which is spatial and directional. Interestingly, it also allows users to decouple

their view camera from their avatar. The purpose of this study was to examine whether deictic reference remained a problem and if so, whether and how people worked around it.

Method

We gathered data in several ways. To facilitate comparison with earlier research we conducted a lab study similar to the approach of Hindmarsh et al. (1998, 2001) and borrowing also from Kraut et al. (2002). Participants logged into *Second Life* in groups of two or three and collaborated on building tasks. Groups were co-present in our lab and arranged so they could hear but not see each other offline. We observed the groups and recorded their screen output and voice conversation for later analysis. We followed up each building session with a focus-group discussion. Each complete session was designed to last an hour and a half, although some enthusiastic groups worked over time, which we allowed.

We observed ten groups; a total of 22 participants. Their ages ranged from 20 to 50. Half were male. Some knew each other before undertaking the task, while other groups were meeting for the first time. We 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 novice use.

After spending a few minutes familiarizing themselves with each other and the lab setup, group members undertook two tasks. The first adapted the 'furniture world' task of Hindmarsh et al. (1998) and the 'jigsaw puzzle' task of Kraut et al. (2002). The researchers provided a set of virtual objects including four walls, two roof parts, two gables, a flag, flagpole and chimney. The group had to assemble these into a house. Some of the objects, such as the flagpole, were unique, meaning they could be referred to by name. Others, such as the walls, were indistinguishable from each other, requiring spatial reference via speech, pointing etc. to identify them. Each group member received a screen-shot of how the house should look when complete.

The second, 'Garden task', used the helper-worker arrangement of Kraut and was designed to force closer collaboration. One group member was designated the helper, while the other(s) were workers. Helpers could not use the building tools themselves, but received a screen shot showing the house now surrounded by extra objects such as garden furniture and a fence, and had to direct their worker(s) to build this scene.

After the building tasks we asked participants about their experience, the problems they encountered, how they solved them, and whether the UI could be enhanced to better support collaboration. Focus-groups were semi-structured to allow exploration of themes.

We used the screen recordings to conduct two quantitative analyses of group performances, counting 'salient events' (Schroeder et al., 2006). We measured the proportion of time users spent with their camera decoupled from their avatar, and we categorized verbal references to objects by their linguistic form. We compared these measures across experience levels and tasks to discover patterns of use.

As a final check we discussed emergent themes with a range of experienced users discovered on the ***SL*** discussion forum and at in-world building classes and competitions.

Findings

We found that participants did indeed encounter communication problems of the kind discovered by Hindmarsh et al. (1998, 2001) and others, but worked around them by experimenting with different referential practices until they were able to convey their meaning. Participants were creative in their use of referential techniques, some of which involved unusual use of avatars and other objects. For example, individuals used their avatar as a kind of public 'cursor' to mark objects, placing it beside objects and saying "this one" or "the one near me". Yet they rarely used pointing, which is afforded by *Second Life* avatars and has been the focus of much research. They "wiggled" objects or temporarily changed the object's color, and created temporary 'marker objects' to mark an object or a location.

Natural language offers several ways to refer to locations. We counted references and categorized them by frame of reference according to the classification of Levinson (1996). References may be 'relative' to the speaker's or receiver's body. Some objects have an 'intrinsic' frames of reference: for example houses may have an obvious front side to which other locations can be compared. 'Absolute' frames involve compass points or prominent objects in the distance. We included a fourth category 'by name or property' to count descriptions such as "the brown thing". We found that groups used all of these forms, in a consistent manner, with reference by name used more than avatar-relative and intrinsic reference, with absolute reference used the least. Voice chat, in combination with visual marking of locations, enabled a reasonably high quality of collaboration even among novices. 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 reasonably well.

The ability to detach camera from avatar allows rapid inspection of different viewpoints, which is useful while building, but makes it impossible to infer with confidence what a player is looking at from their avatar. The results of Hindmarsh et al. suggested this would make some verbal reference impossible to interpret. Yet we discovered that expert players detached their camera from their avatar more than novices, and continued to use avatar-relative reference even though they presumably knew that their partners may not have been viewing the scene through their avatar. Indeed we found that players tended to 'park' their avatars to the side and simply work directly on objects through their movable camera, returning to their avatars only for certain social interactions or to use them as 'cursors' to mark locations. Players made their focus of attention known to others, not through 'embodied' spatial reference such as pointing, but by speech and by directly representing their attention on the objects themselves, for example by moving or tilting objects or changing their color.

Summary

Incorporating voice in large, modern virtual worlds create the possibility of efficient collaboration around virtual objects. While problems of understanding deictic references identified in earlier work persist, it appears that modern virtual world users are able to work around these and get work done. In particular they use their avatars, not so much as embodiments but as cursors to mark locations in space. This suggests that it is talk and activity, not embodiment, which constitutes a player's social presence in virtual space. We concluded that, rather than trying to improve the fidelity of avatar pointing and eye-gaze, a more promising way to assist collaboration in micro space is to provide tools that allow people to speak and to mark out locations directly.

Study 4 represented the end of our explicit examination of the interaction of voice and spatiality. We then proceeded to explore two other important results from our earlier studies - that voice is particularly well-suited to coordinating fast-paced group action, but that it impacts the presentation of self online. To do this we took the opportunity to examine the impact that introducing voice chat had upon a virtual world in which fast-paced action was mostly absent, and role-play was important.

2.5 Study 5: Voice in a pseudonymous social world

Studies 1 and 2 showed that the rich social presence (intense cues about player identity and state) conveyed by voice introduced problems for players such as diminished ability to role-play, reluctance to communicate with strangers, and increased vulnerability to abuse. If the acceptability of voice in online games relies on a delicate balance between the benefits of team coordination and the problems of controlling social presence, how might voice influence the user experience of non-game-oriented virtual worlds? We explored this by investigating the introduction of voice into *Second Life*, a 'social' virtual world whose many dedicated users engage in a diverse set of activities that includes content creation, avatar customization, role-play, education, commerce and socializing. This study was conducted in parallel with study 4.

When this study began there was an emphasis on pseudonymity in *SL* and exposure of offline identity was proscribed. *SL*'s vendor Linden Lab announced in 2007 that it would introduce a voice channel later that year. This sparked a storm of controversy among users (Boellstorff 2008; pp. 113-4). Echoing Bartle (2003), 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 *SL* if *anyone* used voice. In 2009, two years after the introduction of voice, a Linden Lab press release stated that on the one hand, *Second Life* had become a major player in international telephony and carried a significant proportion of the world's long-distance voice conversations; yet only 50% of *SL* users had adopted the voice channel. Study 5 was aimed at shedding light on this conundrum. Why did so many *SL* users reject voice while so many others embraced 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 the virtual world?

Method

To study attitudes and practices related to voice communication in *SL* we interviewed users, convened in-world discussions, analyzed forums and blogs, and conducted participant research into both recreational and workplace use of *SL* from the introduction of voice in 2007 through to the end of 2009.

We interviewed 7 experienced users from a variety of backgrounds who used *SL* for activities including socializing, teaching, business and art. All lived in Australia. Three were male and all were aged between 20 and 40. Interviews were open-ended and lasted one to two hours. Some interviews took place in *SL*, using voice or text according to the interviewee's preference, while some were face-to-face. We also moderated a focus-group about voice within a long-running weekly *SL* discussion group. This lasted an hour and comprised about twenty participants, including long-term *SL* users and commentators. We monitored the official *SL* user forum, two popular email-lists, and several prominent blogs during the research period, saving relevant threads for later analysis.

One of the researchers were active users of *SL* and other VWs throughout the research period and was involved in several working groups which held regular meetings in *SL*. These included a geographically-dispersed research group within a large institution, and a conference organizing committee. This required that we deal with problems of voice and text communication in our own work practice. We organized a weekly in-world research seminar series that utilized both voice and text. One author took part in several mixed-reality music performances that used *SL*. This period of intensive participant research allowed us to develop insights and cross-check experiences reported by informants.

Using grounded methods we coded data into themes which reflect the concerns of our participants. These are presented in the next section.

Findings

Players were more ambivalent about voice in *Second Life* than in the games we studied earlier. Its introduction 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"*. Others some felt that extreme pro- or anti-voice positions were inappropriate and that players should be left to choose the medium they preferred. One interviewee 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 players: suggestions included DJs and live musicians, people helping newbies, trolls, and those engaged in online sex. Informants who used *SL* for business meetings, telecommuting, education and drama highly valued voice. Some players praised the ability of voice to convey more richness and subtlety of meaning, more quickly than when typing. However as in earlier studies, informants felt that voice conveyed more information about the identity of the person speaking than did text – such as their gender, ethnicity and age – interfering with the anonymity which had previously characterized the virtual world experience and which

many players 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* players for rejecting voice. Some felt that voice had robbed *Second Life* of its "atmosphere", referring to the "crassness" of voice and "interruptions by real humans". Some referred to the distinction, popular among *SL* players, between immersionists (role-players who sought an alternative reality in the virtual world) and augmentationists (who used the VW to communicate with people known to them offline) (Boellstorff 2008). Immersionists complained that adding voice simply gave augmentationists "a better telephone to chat with their real-life friends". Some felt that players' opinions about voice was determined by personality, and that a text-only world was a haven for shy introverts which would be shattered by the introduction of voice: *"If you introduce sound the whole ambience changes: the shy are revealed as shy, and the noisy start to dominate."*

Voice had a similar impact upon role-playing to that reported in study two, with people who played with opposite gendered avatars being particularly apprehensive about voice. Males playing female avatars related being "*out-ed*" by adopting voice, and feeling so ashamed as to leave online groups. Females playing male avatars feared the "*feeding frenzy*" that would ensue if they were exposed: though one conducting a business in *SL* openly as a woman reported being treated better by clients. Furthermore, since reluctance to use voice was seen as a sign of reluctance to expose identity, people who disliked voice for other reasons feared being suspected of having something to hide: *"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."*

While concepts like 'immersion' tend to convey the idea of a player being completely absorbed in their current activity in a virtual world, it appeared in fact that many *SL* players multi-tasked: either combining *SL* use with offline activities, or by engaging simultaneously in multiple conversations within *SL*. Informants felt that text was better suited to this: *"especially when you're multitasking as I often am, text is easier."* A related activity is holding multiple simultaneous conversations, which players found easier in text: *"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."* Many reported simultaneous use of voice and text, for example in public meetings and classrooms in which the speaker used voice and the audience typed questions. But some liked to save text transcripts of discussions, which could not be done in voice, and found that discussions involving many players worked better in text. Usability was an advantage of voice as in earlier studies, though some pointed out that some disabilities made text preferable (Carr 2010 discusses the complex reactions some hearing-impaired *Second Life* users experienced upon the introduction of voice).

Second Life's spatial voice configuration received some criticism. Whereas in physical space people have an intuition about how far their voice transmits and thus who can hear them, the relationship between volume and distance was not obvious to *SL* players (this problem was noted also by Erickson et al. 2011). An interviewee was

concerned that they couldn't control transmission distance by changing the volume with which they spoke. Mimicking the physics of physical-word voice transmission also reintroduced physical-world problems to virtual space: one player 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.

As in studies 1 and 2, players complained about unintended sounds being broadcast into the virtual world: *"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".* Some informants reported maintaining awareness of which sounds in their environment they might unintentionally transmit. Some were concerned that this 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 physical galleries and music venues: audience members chatting and disturbing the experience of fellow attendees. Players were also wary of sound from the virtual world – or their own voice – being heard by those around them: *"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."* The uncertain broadcast radius, combined with players' ability to detach eyes and ears from avatar without this being visible to others, also opened the possibility of eavesdropping in *virtual* space (also noted by Irani et al. 2008).

Players felt that griefing was worse in voice than in text, with one participant claiming that voice was "*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."* Some felt that an insult *"becomes more personal when voice is used".* Some thought voice attracted people who were disposed to griefing, while some felt voice exaggerated offline communication problems. Griefers did not only broadcast their voices but music and bodily and other offensive noises. Some players asked for an etiquette of online voice to be developed. One participant noted that: *"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."* Through making certain players 'invisible', virtual space became narrowly cultured, with the voice channel exacerbating the hostility expressed towards present-yet-invisible players.

Summary

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. The introduction of voice into this popular role-playing world provoked passionate responses (for and against) among very

many users. Voice changed the *Second Life's* atmosphere, the kinds of people that were attracted to use it, and the kinds of activities they carried out there.

For our most recent study of voice in virtual worlds we have evaluated a new game which uniquely combines features of the FPS and MMORPG genres while using spatial voice to enable new forms of player interaction and an unprecedented degree of social realism.

2.6 Study 6: Spatial voice and moral dilemmas in *DayZ*

Though we studied a form of spatial voice as early as 2006, and a spatial system was implemented in *Second Life* in 2007, implementation of spatial voice in the first-person shooter genre remains rare. However *DayZ*, a zombie survival themed sandbox Massively Multiplayer Online First-Person Shooter (MMOFPS), was a popular 2012 release which offered voice communication that employed the 'sound in air' metaphor. In the *DayZ* virtual world, if a player's virtual avatar is within 50 metres of another player's avatar, they can engage in voice chat (which is stereo-directional, thus also allowing players to discern each other's locations).

DayZ is notable for a number of other reasons. First, unlike most FPS games, in which death is a mere inconvenience, *DayZ* implements 'character-death' whereby a player's entire in-game advancement is lost (Carter et al. 2013; Heaven, 2013). Since players begin the game with few items and must scavenge the food, water, medication, ammunition and firearms necessary to safely navigate the environment, losing all possessions is a significant negative consequence. Secondly, players are not configured into formal teams but are able to kill any player they encounter and can loot the corpses of dead players to appropriate the fruits of their labor. Yet at the same time cooperation between players is sometimes necessary in this harsh environment, to achieve game goals and in order to enjoy social experiences during game play. This leads to moral dilemmas for players, who must constantly decide whether to "cooperate or defect" as this choice is labeled in the Prisoner's Dilemma model. By allowing any player to speak with any other whom they meet in the game, spatial voice makes possible an ongoing negotiation of this moral dilemma.

Method

We studied social interaction and game play in *DayZ* in 2012, when it was released as a 'mod' to the FPS game *Arma 2*, and again in 2013 when it was re-released as a standalone game. We undertook participant observation and a comprehensive review of player-generated videos uploaded to sites such as *YouTube*, stories of game play recounted on forums like *Reddit*, and lengthy online debates about *DayZ*'s ongoing design. Videos were categorized according to the types of voice-based interactions that they depicted. We analyzed them to understand how the unique affordances of spatial voice in an all-to-all combat game affected the social experience of play. As a result of the self-selection of these videos (with only interesting ones being uploaded), these primarily represented 'interesting' or unusual interactions, which we felt was acceptable as we wanted to understand the different and unique interactions that *DayZ* afforded in

comparison to our earlier studies. Similar coding was performed on accounts of play on online forums where we sought to identify the full breadth of interactions that occur in *DayZ*. Over 60 videos and stories were analysed, at which point we felt we had reached data saturation.

We also interviewed 8 *DayZ* players about their experiences, in particular using these discussions to extrapolate the experience of players during these interactions, as well as their style and affordances. In addition, reflecting the auto-ethnographic inspired approach of authors like Linderoth et al (2012), we also recorded play sessions in order to note the use of voice by participants and those they encountered. All of these players were male, aged between 21 and 28, and had significant experience playing online games and using voice communication software.

Findings

Proximity-based voice chat enables a unique style of player interaction in *DayZ*. Most FPS games involve clear demarcations of 'friends' and 'enemies'. But spatial voice allows *DayZ* players to speak with strangers encountered in the game and engage in interaction which may be collaborative, treacherous, result in mutually beneficial trades, or involve a simple mundane greeting or interaction in passing.

The zombie-infested environment of *DayZ* is dangerous to traverse. Collaboration between players significantly reduces this risk. Consequently players are motivated to team up and work together, and they are able to do this with strangers met through voice chat. However potential cooperators must be on guard against the possibility of treachery. *"I spotted a flare in a town and asked if the thrower was friendly, they announced yes and asked to team up. I said yes on the agreement they show themselves as a sign of trust. They did so and I then approached. Once contact had been made we then proceeded to explore the town"*. Voice, as well as non-verbal and proxemic interactions, such as lowering a weapon, or standing at a reasonable distance if wielding a fire axe, served to enable these ad-hoc peaceful collaborations.

Some interactions can also be altruistic, as reported by one interview participant: "*We were reasonably well armed and watched a newly spawned player fall off a bridge and break his legs. With 2 snipers watching over us, myself and a friend travelled down there to help him. Using direct chat we identified ourselves as friendly and that we would fix his legs and heal him. Once we had done that we made it clear that while we were helping him - he was not to follow us as we retreated and that if he did we would return fire on him."* Were voice restricted to channels or teams, game-play like this would not be possible.

At the same time spatial voice allows treacherous play; the use of voice to deceive and betray (Carter 2014). One participant related one such experience;*"We had some awesome stuff, 2 pistols, heaps of meds, then we saw a dude with no weapons. He asked if we could give him a blood transplant so we figured we would, then as we were picking up the blood bag he went into our pack, took our stored pistol and shot us..... MOTHER*

*F*****R!!!!"*. Thus the ambiguity of social relationships was exploited tactically in a fashion not dissimilar to the deception seen in study 3.

Other players exploited the directionality of spatial voice. In one group play session, two participants were killed when responding to a friendly invitation; they explained that by responding, they gave away their location (a building promptly attacked by grenade). Another participant reported a situation where another player gave away their location by asking "*friendly?*", alerting the player to the presence of this other player; *"I said no words even when he got to the top of the stairs and i shot him in the head from behind.*" The need for cooperation, combined with the danger of treachery and the drastic consequence of death, contributes to the excitement of *DayZ.* The consequential emotional intensity of play was frequently reported as an attraction to play this game.

Proximity-voice also impacts the player experience in mundane, non-confrontational ways. One participant reported; *"sneaking through a town I heard a couple of people talking to each other in the supermarket, that was interesting, hearing them while they think they are alone."* Another, upon overhearing two players conversing, used proximity-voice while hidden to confirm that they had passed, without having to risk reveal their own location or killing players unnecessarily. Where collaboration is not invited, proximity-voice is also used by players to negotiate a peaceful passing rather than ending in the death of either player (with its loss of significant advancement). Whereas in earlier studies players who did not use voice became 'invisible', in *DayZ*, non-participation in voice-chat was sometimes interpreted as an act of hostility; mute players soon became dead players.

Voice chat is often used creatively by *DayZ* players to add to the experience of other players (and possibly as a form of role-play); "*The first interaction I had via direct voice chat was someone recreating the noises from the Friday the 13th series of films, shortly before they shot me*." A popular *YouTube* video shows a player on a bicycle using voice to play the hit song "Ridin' Dirty" as they ride past players in the virtual world. Other videos depict players using voice to trick and 'grief' players. We also note that the configuration of voice in *DayZ* was beneficial for combating in-game harassment. While in other FPS and MMORPG games female players are placed by the system into channels or lobbies (with the only other alternative being to not participate in voice chat), the all-versus-all combat of *DayZ* meant that female players could kill offending players or move away from them. Despite this affordance, less than 1% of 1800 respondents to an online survey we recently deployed (not included in this analysis) identified as female.

While most people seem to play *DayZ* as individuals and team up opportunistically, some groups of players known to each other have formed ongoing cooperative teams reminiscent of MMORPG guilds. These rely heavily on third-party 'radio' voice software (such as TeamSpeak or Ventrilo) to overcome the problem of coordinating widely-distributed movement through space. However some players claim that using radio detracts from the authentic *DayZ* experience, and call for spatial voice only servers. Thus as in study 3, conflict emerged over the trade-off between realism and utility, with participants using push-to-talk to communicate with permanent team-members and ad-hoc groups. This sometimes caused confusion \when players in ongoing cooperative

teams were also co-located in the virtual world, with some continuing to use third-party programs while others switched to spatial voice, creating the appearance of someone speaking to themselves.

Summary

Spatial voice has rarely been implemented in commercial games: an exception is *Second Life*. Combat games have invariably offered team-based radio voice. Some participants in our MMORPG studies speculated whether spatial voice might allow players to speak with strangers they encountered while roaming the virtual world. *DayZ* has demonstrated that implementing spatial voice in an MMOFPS opens up fascinating new forms of gameplay by allowing players to communicate with whomever they encounter in their travels, with the rapidity and emotional expression that voice allows. This enables an exciting, lawless world of moral dilemmas. This has introduced a new level of *social* realism to a genre normally concerned with graphical and tactical realism.

3. DISCUSSION

In this final section we draw out the key themes that arose in our research, highlighting issues that arise across varieties of virtual world and voice configurations, and the lessons that can be drawn from these findings.

3.1 Voice transforms the experience of virtual worlds

Our studies of *Xbox Live*, MMORPGs and *Second Life* in particular showed that communicating by voice radically transforms the player experience of online games and virtual worlds. Voice breaches whatever boundary exists between virtual and physical worlds (cf. Taylor 2006, Lehdonvirta 2010). It projects more information about the identity of players than text. This makes the play experience more sociable for some, and can transform play, but can interfere with role-play and pseudonymity. The existence of different sub-groups of users with different attitudes to the use of virtual worlds adds complexity: our research showed how these attitudes map to people's choice and use of media.

3.2 Advantages and challenges of voice

According to Clark (1996), face-to-face conversation in small groups is the primal form of language. However, voice can be disadvantageous in virtual worlds. Our MMORPG study emphasized that voice is ephemeral, leaving no trace of utterances, and does not scale well to large groups. Our *Xbox Live*, MMORPG and *Second Life* studies showed that voice more susceptible to abuse and is not preferred by the shy nor by those seeking to control self-presentation. Multiple studies emphasized that voice is more prone to being received by unintended recipients and does not suit the multi-tasker. Voice channels convey background noises and household sounds, which can impact privacy and immersion. Virtual worlds are (for most) not workplaces, and the advantages that speech confers at work do not necessarily translate to play and especially to role-play.

3.3 Media richness and virtual worlds

Media richness theory (MRT) (Daft et al, 1987) proposed that in the workplace, spoken conversation was superior to written forms of communication in terms of efficiency, symbol variety, and delivery of multiple cues and emotional tones. Critiques of media richness theory have demonstrated 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. Our XBL, MMORPG and *SL* studies 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. We 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 player'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) found that educators' intention to use Second Life was predicated on its offer of rich communication.

However our findings emphasize that the key outcome of post-MRT research – that richer doesn't necessarily mean better and that face-to-face is not a gold standard to which all mediated communication should aspire – applies to virtual worlds as much it does to other communication contexts. It is unlikely that game players want a maximally-rich medium. Even enthusiastic voice users playing MMORPGs recognize a range of problems related to unwanted projection of social presence.

This view accords with Golub's (2010) critique of the notion 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 players of beautifully-rendered MMORPGs are willing to use mods which replace the 3D scene with numbers and charts that explicitly represent the underlying game state, in order to gain advantage in the game. Some players also turn the game's immersive sound effects off so that they can concentrate on game-play and on each other's conversation.

Our MMORPG and *Second Life* studies in particular highlighted that people like to multi-task while they are communicating with non-collocated others. 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 player into one task. Immersion and flow can be desirable in entertainment contexts (Badique et al., 2002) but may be 'a bug rather than a feature' during instrumental use. Perhaps this underlies the failure of virtual worlds in workplace contexts compared to 2D, text-based technologies such as the web.

3.4 Fictional 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 this research and a consistent finding in prior work. Consider the excitement that Nardi (2010) reported experiencing during her first *WoW* session: *"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."* (Nardi, 2010, p. 5)

Our studies show that while game players enjoy the presence of others, the identity they wish to project to others is not necessarily a straightforward extension of offline identity. Many VW users want to be in the world, but do not want to be themselves in it, and do not want reality intruding in the form of trash-talk and background noise. They want an identity that persists over time and yet is constructed, and they put significant effort into maintaining it. We use the phrase *fictional social presence* to describe this persistent, yet fictionalized, identity that some players wish to project into virtual worlds. It describes the paradox of role-play: that players desire social interaction but are constructing their identities in a fictional setting and may wish to protect offline identities. Using voice makes this a more difficult enterprise.

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. Thus the use of voice changes the social opacity of virtual worlds, with implications for the construction and maintenance of fictional social presence.

3.5 Multi-tasking and multi-contexting

Performance construction is particularly apparent when VW players exploit lean media to multi-task and even 'multi-context', engaging in impression management and multiple conversations with a range of people both online and physically collocated (cf. Jones, 2002). Dourish (2004) and others have 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 in offline contexts, people are aware of who in their physical context might be able to overhear them, and tailor communication to suit (Madell and Muncer, 2007).

Multi-tasking need not imply deception of a malicious kind. VW users may feel themselves to be present with collaborators, yet at the same time maintain an awareness of what is and is not being transmitted about activity in their physical context. They use this awareness to engage in mundane activities such as answering the phone, talking to others who are co-present, or using other software. Such activities need not be accountable to fellow players for effective cooperation 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. 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 players are so absorbed in the simulation that they forget their physical surroundings (Biocca et al., 2003). However an employee using a

VW for communication at the office, a parent engaging simultaneously in family life and *Second Life*, or a teenager playing a game while chatting with collocated friends, cannot afford to be so immersed in the virtual world that they ignore the physical one.

Spatial voice transmission is designed to provide an intuitive way to manage conversation contexts: players can space their avatars in order to control who hears whom. It also enables voice-chat with strangers encountered during travel (study 6). However the distance over which voice travels is often opaque to players (Erickson et al. 2011 and our study 3), and the movable camera in *Second Life* makes judging audibility impossible (Irani 2008 and our study 4). During conversation, players need to continually calculate who can overhear them. This situation is complex to manage and restricts the utility of voice.

3.6 Simulated telecommunication in simulated space

The systems we studied simulated a space and ways of delivering messages within the space. There is a virtual distance between any two players, 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. By analogy with traditional communication media we consider these message-delivery mechanisms to be 'virtual media'. This term highlights the fact that virtual worlds typically emulate existing media that are in use outside the VW context. While advances in hardware and bandwidth are opening new possibilities for voice channel design, in over a decade designers have rarely strayed from two standard configurations: the walkie-talkie channel in which team members hear each other equally but do not hear opposing teams, and spatial voice which propagates like physical sound.

Other metaphors are possible, and designers could consider which are appropriate to particular game-worlds (for example medieval warriors did not use radios). Potential metaphors which have been little explored include mobile and fixed-line telephones: these would suit games set in particular historical eras. Landline would require the player to be situated at a nominated point in the game world to use voice. Virtual mobile phones might restrict communication to pairs of players and require them to know phone numbers. The ability to communicate might need to be earned, just as weapons are typically awarded to players upon achieving particular game goals[2].

3.7 Technical difficulties

Technical glitches are mundane problems. Yet these were reported as significant by voice users in all of our studies. Voice has inherent complexities relative to text. The path of sound from sender's mouth to receiver's ear involves many more steps than that of text from fingers to screen. Voice is an analogue signal that must first reach a microphone, which players may not position well, while those without a headset may use poor-quality

[2] This is currently being explored in the ongoing development of *DayZ*, where 'walkie-talkies' can be switched on and left on the ground, allowing players to listen in on other players when not nearby, or project their voice to humorous or strategic effect. Unfortunately, this feature was introduced after our data collection had completed.

microphones built into laptops. The distance from mouth to microphone can change unintentionally, affecting volume. These variables make volumes differ at random; yet for a conversation to sound natural, the voices one is hearing should be at about the same volume except where change is intentional or due to distance. There is no analogy to volume in a text channel (other than the general understanding that using 'all caps' is akin to shouting). Digitization quality varies from computer to computer, and digital sound consumes more bandwidth than does text. Network latency is more likely to create problems in turn-taking and comprehension when speaking than typing. All 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, as seen in study 2.

The systems we studied offered players no feedback as to whether they were transmitting too loudly or too softly, if their voice was distorted or garbled, or if they were transmitting background noise or other unwanted noises such as the sound of their breath. Text on the other hand is perceived as reliable and accessible, and is used as a backup 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 a virtual world.

3.8 Implications for design

We hesitate to offer design suggestions because there are probably no "silver bullet" solutions to the problems our work has uncovered. The use of voice chat in virtual worlds may always involve trading off the advantages of easy, fast communication against the disadvantages of channel congestion, lack of a transcript, awkwardness with strangers and diminished opportunities for role-play. Radio users encounter a similar array of advantages and difficulties (Camp et al. 2000, Woodruff and Aoki 2003, Toups et al. 2011). Some technology vendors have offered a potential workaround to some of these problems in the form of voice-disguising software (some of our participants proposed this also, without having tried it). An example is *Second Life's* recent "voice morphing" feature. However anecdotal evidence suggests that use of this is treated with some suspicion and that it has had limited uptake and is not desired by "augmentationists" anyway.

Perhaps the best outcome would be for not just designers but users to recognize the complex set of benefits and drawbacks that voice chat brings to games and how these vary with situation, task and personality. Rather than seek a technological fix for voice, it might be more productive that designers and players do not demand that individuals use voice, or that it at least be recognized where voice is problematic and why some people will prefer not to use it. If indeed we were to conclude that a deeper understanding of these issues among the gamer community is the best outcome one can hope for, it would support the view that the value in studying technology use need not necessarily lie in generating design suggestions (Dourish 2006).

3.9 Conclusion

Through a series of studies we have developed an understanding of how the availability and configuration of voice in virtual worlds influences play and player experience. Our work connects game studies to HCI and the study of CMC. Simply adding all-to-all voice to a virtual world and 'hoping for the best' is unlikely to lead to a successful implementation: users may reject the channel and even the virtual world itself. We conclude that voice radically transforms the experience of online gaming, making virtual spaces more intensely social but making them into maelstroms of impression management, identity play, and ambiguity over what is being transmitted to whom.

Virtual world use involves people in physical settings projecting constructed identities into fictional spaces. Player-to-player communication lies at the intersection between fantasy and reality. The suitability of particular media to particular situations depends on a dynamic compromise between the need for people to understand collaborators and their desire for privacy and identity-play.

REFERENCES

Aarsand, P. & Aronsson, K. (2008). Response cries and other gaming moves—Building intersubjectivity in gaming. *Journal of Pragmatics* 41 (8). 1557–1575.

Aarseth, E, (2008). Allegories of space: the question of spatiality in computer games. In F. von Borries et al. (Eds.), *Space Time Play: computer games, architecture and urbanism: the next level* (pp. 44-55) . Basel: Birkhauser.

Badiqué, E., Cavazza, M., Klinker, G., Mair, G., Sweeney, T., Thalmann, D. & Thalmann, N. (2002). Entertainment applications of virtual environments. In Stanney, K. (Ed.), *Handbook of virtual environments* (pp. 1143–1166). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Bailenson, J. N., & Blascovich, J. (2004). Avatars. In Bainbridge, W. S. (Ed). *Berkshire encyclopedia of human-computer interaction (Vol. 2).* Berkshire Publishing Group.

Bartle, R. (2003) Not Yet You Fools! Article posted at Game Girl Advance website. Retrieved 7th October 2011 from www.gamegirladvance.com/2003/07/not-yet-you-fools.html.

Biocca, F., Harms, C. & Burgoon, J.K. (2003). Toward a more robust theory and measure of social presence: Review and suggested criteria. *Presence* 12(5). 456-480.

Blinka, L. (2008). The Relationship of Players to Their Avatars in MMORPGs: Differences between Adolescents, Emerging Adults and Adults. Cyberpsychology: *Journal of Psychosocial Research on Cyberspace* 2(1), article 5.

Boellstorff, T. (2008). *Coming of Age in Second Life*. Princeton: Princeton University Press.

Boustead, P., Safaei, F. & Dowlatshahi, M. (2005). DICE: Internet delivery of immersive voice communication for crowded virtual spaces. In *Proceedings of the 2005 IEEE Conference on Virtual Reality*. 35-41. IEEE Computer Society.

Bowers, J., Pycock, J. & O'Brien, J. (1996). Talk and embodiment in collaborative virtual environments. In *Proceedings of the SIGCHI conference on Human factors in computing systems: common ground*. New York: ACM.

Brown, B. & Bell, M. (2004). CSCW at play: 'There' as a collaborative virtual environment. In CSCW '04 *Proceedings of the 2004 ACM conference on Computer supported cooperative work*. New York: ACM.

Camp, P. J., Hudson, J. M., Keldorph, R. B., Lewis, S., & Mynatt, E. D. (2000). Supporting communication and collaboration practices in safety-critical situations. In *CHI'00 extended abstracts on Human factors in computing systems* (pp. 249-250). ACM.

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

Carr, D. (2010). Constructing disability in online worlds: conceptualising disability in online research. *London Review of Education*, 8(1), 51-61.

Carter, M., Gibbs, M., Wadley, G. (2012a). Avatars, Characters, Players and Users: Multiple Identities at/in Play. In *Proceedings of the 24th Australian Computer-Human Interaction Conference* (ozCHI'12). New York, ACM Press. 68-71.

Carter, M., Wadley, G., Gibbs, M. (2012b). "Friendly, Don't Shoot!": How Communication Design can enable Novel Social Interactions. In *Proceedings of the 24th Australian Computer-Human Interaction Conference* (ozCHI'12). New York, ACM Press. 72-75.

Carter, M., Gibbs, M., Wadley, G. (2013). Death and Dying in DayZ In *Proceedings of Interactive Entertianment 2013*, Melbourne

Carter, M. (2014). Massively Multiplayer Dark Play: Treacherous Play in EVE Online. In T. Mortensen & J. Linderoth (Eds.), *The Dark Side of Game Play.* London: Routledge.

Chen, M. (2012) *Leet noobs: the life and death of an expert player group in World of Warcraft*. Peter Lang: New York.

Churchill, E. F., Snowdon, D. N. & Munro, A. J. (Eds.) (2001). Collaborative virtual environments: digital places and spaces for interaction. London: Springer.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

Daft, R. L., Lengel, R. H. & Trevino, L. K. (1987). Message equivocality, media selection, and manager performance: Implications for information systems. *MIS Quarterly* 11(3) pp. 355- 366

Dourish, P. (2004). What we talk about when we talk about context. *Personal and Ubiquitous Computing* 8(1). 19–30.

Dourish, P. (2006). Implications for design. In CHI '06 Proceedings of the SIGCHI conference on Human Factors in computing systems. ACM: New York.

Ducheneaut N. & Moore, R. 2004. The social side of gaming: a study of interaction patterns in a massively multiplayer online game. In Proceedings of the 2004 ACM conference on Computer supported cooperative work (CSCW '04). ACM, New York, NY, USA, 360-369.

Ducheneaut, N., Yee, N., Nickell, E. & Moore, R. (2006). "Alone together?": Exploring the social dynamics of massively multiplayer online games. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. New York: ACM.

Ducheneaut, N., Moore, R. & Nickell, E. (2007). Virtual "Third Places": A Case Study of Sociability in Massively Multiplayer Games. *Computer Supported Cooperative Work* 16(1-2), 129-166.

Ducheneaut, N., Wen, M-H, Yee, N. & Wadley, G. (2009). Body and mind: a study of avatar personalization in three virtual worlds. In *Proceedings of the 27th international conference on Human factors in computing systems*. New York: ACM.

Erickson, T. & Kellogg, W. (2000) Social translucence: An approach to designing systems that support social processes. *ACM Transactions on Computer-Human Interaction* 7(1). 59-83.

Erickson, T. Shami, S., Kellogg, W. & Levine, D. (2011). Synchronous interaction among hundreds: an evaluation of a conference in an avatar-based virtual environment. In *Proceedings of the 2011 annual conference on Human factors in computing systems*. ACM: New York.

Gibbs M.R., Hew, K. and Wadley G. (2004) Social Translucence of the Xbox Live Voice Channel. In M. Rauterberg (ed.) *Proceedings of ICEC 2004: IFIP 3rd International Conference on Entertainment Computing*, Eindhoven, Netherlands, 1-3 Sept. Springer-Verlag Berlin pp. 377-385.

Gibbs M.R., Wadley, G. and Benda P. (2006) Proximity-Based Chat in a First Person Shooter: Using A Novel Voice Communication System for Online Play. *In Proceedings of IE 2006: Third Australasian Conference on Interactive Entertainment*. Perth, Australia: Murdoch University, 96-102.

Glaser, B. G. & Strauss, A. L. (1967) *The discovery of grounded theory; strategies for qualitative research*. Chicago: Aldine.

Goffman, E. (1959). *The Presentation of Self in Everyday Life*. New York: Anchor.

Golub, A. (2010). Being in the World (of Warcraft): Raiding, Realism, and Knowledge Production in a Massively Multiplayer Online Game. *Anthropological Quarterly* 83(1). 17–46.

Gray, K. (2013). Collective Organizing, Individual Resistance, or Asshole Griefers? An Ethnographic Analysis of Woman of Color in Xbox Live. *Ada: A Journal of Gender, New Media and Technology* (2). http://adanewmedia.org/2013/06/issue2-gray/

Gutwin, C. & Greenberg, S. (2002). A Descriptive Framework of Workspace Awareness for Real-Time Groupware. *Computer Supported Cooperative Work* 11(3-5). 411-446.

Halloran, J., Fitzpatrick, J., Rogers, Y. & Marshall, P. (2004). Does it matter if you don't know who's talking? Multiplayer gaming with voiceover IP. In *CHI '04 extended abstracts on Human Factors in Computing Systems*. New York: ACM.

Harrison, S., Tatar, D. & Sengers, P. (2007). The three paradigms of HCI. In Proceedings of CHI 2007 (alt.chi), San Jose, CA, May 2007.

Heaven, D. (2013) The rise of consequences in video games. In New Scientist blog. Available at http://www.newscientist.com/article/dn24325-buttonmasher-the-rise-of-consequences-in-video-games.html#.UlodJBBvDfP .

Hindmarsh, J., Fraser, M., Heath, C., Benford, S. & Greenhalgh, C. (1998). Fragmented Interaction: Establishing Mutual Orientation in Virtual Environments. In *Proceedings of the 1998 ACM conference on Computer Supported Cooperative Work*. New York: ACM.

Hindmarsh, J., Fraser, M., Heath, C. and Benford, S. (2001) 'Virtually missing the point: configuring CVEs for object-focused interaction', in Churchill, E. F., Snowdon, D. N. and Munro, A. J. (eds.): Collaborative Virtual Environments: Digital places and spaces for interaction, Springer, London, pp. 115-139.

Irani, L. C., Hayes, G. R. & Dourish, P. (2008). Situated practices of looking: visual practice in an online world. In *Proceedings of the 2008 ACM conference on Computer Supported Cooperative Work*. New York: ACM.

Jones, R. H. (2002). The problem of context in computer mediated communication. In Levine, P. & Scollon, R. (Eds.). Discourse and Technology: Multimodal discourse analysis. Washington: Georgetown University Press. 20-33.

Kraut, R. K., Gergle, D. & Fussell, S. R. (2002). The Use of Visual Information in Shared Visual Spaces: Informing the Development of Virtual Co-Presence. In *Proceedings of the 2002 ACM conference on Computer Supported Cooperative Work*. New York: ACM.

Kuznekoff, J. H., Rose, L. M. (2012). Communication in Mutliplayer Gaming: Examining player responses to gender cues. *New Media Society* 15: 541-556.

Landgren, J. (2006). Making action visible in time-critical work. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 201-210). ACM.

Lehdonvirta, V. (2010). Virtual Worlds Don't Exist: Questioning the Dichotomous Approach in MMO Studies. *Game Studies* 10(1).

Levinson, S. C. (1996). Frames of reference and Molyneux's question: cross-linguistic evidence. In Bloom, P., Peterson, M.A., Nadel, L. & Garrett, M.F. (Eds.). *Language and Space* (pp. 109-170). Cambridge: MIT Press.

Linderoth, J., Björk, S. & Olsson, C. (2012). Should I Stay or Should I go? Boundary Maintaining Mechanisms in Left 4 Dead 2. In *Proceedings of DiGRA Nordic 2012*. http://www.digra.org/wp-content/uploads/digital-library/12168.12472.pdf

MacCallum-Stewart & Parsler (2008). Role-play vs. gameplay: The difficulties of playing a role in World of Warcraft. In Corneliussen, H. & Rettberg, J. (Eds.). *Digital Culture, Play, and Identity: A World of Warcraft Reader*. Cambridge: MIT Press. 225-246.

Madell, D. E. & Muncer, S. J. (2007). Control over Social Interactions: An Important Reason for Young People's Use of the Internet and Mobile Phones for Communication?. *Cyberpsychology and Behavior* 10(1). 137-40.

Milik, O. (2014). The Digital Grind: Time and Labor as Resources of War in EVE Online. In M. Carter, K. Bergstrom & D. Woodford (Eds.), *Internet Spacehsips are Serious Business*. Minneapolis: The University of Minnesota Press.

Moore, R., Ducheneaut, N. & Nickell, E. (2007). Doing virtually nothing: Awareness and accountability in massively multiuser online worlds. *Computer Supported Cooperative Work* 16(3). 265-305.

Nardi, B. & Harris, J. (2006). Strangers and friends: Collaborative play in World of Warcraft. In *Proceedings of the 2006 20th anniversary conference on Computer Supported Cooperative Work*. New York: ACM.

Nardi, B. A. (2010). *My life as a night elf priest: an anthropological account of World of Warcraft*. Ann Arbor: University of Michigan Press.

Neustaedter, C. & Fedorovskaya , E. (2009). Capturing and Sharing Memories in a Virtual World. In *Proceedings of the 27th international conference on Human factors in computing systems*. New York: ACM.

Nilsson, A., Heldal, I., Axelsson, A-S. & Schroeder, R. (2002). The long-term uses of shared virtual environments: an exploratory study. In Schroeder, R. (Ed.) *The Social Life of Avatars*. London: Springer. 112-126.

Nitsche, M. (2008). *Video game spaces: image, play, and structure in 3D game worlds*. Cambridge: MIT Press.

O'Malley, C., Draper, S., et al. (1984): Constructive Interaction: A Method for Studying Human-Computer Interaction. In: Shackel, B. (ed.): Proceedings of Interact 84, North Holland, Amsterdam, 269-274

Preece, J. & Maloney-Krichmar, D. (2003). Online communities: Focusing on sociability and usability. In Jacko, J. A. & Sears, A. (Eds), The Human-Computer Interaction Handbook Mahwah, NJ: Lawrence Erlbaum. 596-620.

Raybourn, E. M. (2001). Designing an emergent culture of negotiation in collaborative virtual communities: the DomeCity MOO simulation. In Churchill, E. F., Snowdon, D. N., & Munro, A. J. (Eds.), *Collaborative virtual environments: digital places and spaces for interaction*. London: Springer.

Roussel, N. & Gueddana, S. (2007). Beyond "beyond being there": towards multiscale communication systems. In *Proceedings of the 15th international conference on Multimedia*. New York: ACM.

Saeed, N., Yang, Y. & Sinnappan, S. (2008). Media richness and user acceptance of Second Life. In *Proceedings of Ascilite 2008*, Melbourne.

Sallnas, E-L. (2002). Collaboration in multi-modal virtual worlds: comparing touch, text, voice and video. In Schroeder, R. (Ed.), *The social life of avatars: presence and interaction in shared virtual environments* (pp. 172-187). London: Springer.

Schiano, D. (1999). Lessons from LambdaMOO: A social, text-based virtual environment. *Presence* 8(2). 127-139.

Schmidt, K. (2002). The problem with `awareness': Introductory remarks on `awareness in CSCW'. *Computer Supported Cooperative Work* 11 (3-4). 285-298.

Schroeder, R. (2002) (Ed.). The social life of avatars: presence and interaction in shared virtual environments. London: Springer.

Schroeder, R. & Axelsson, A.-S. (Eds.) (2006). Avatars at work and play: collaboration and interaction in shared virtual environments. Dordrecht: Springer.

Schroeder, R., Heldal, I. and Tromp, J. (2006) 'The Usability of Collaborative Virtual Environments and Methods for the Analysis of Interaction' Presence: Teleoperators & Virtual Environments, vol. 15, no. 6, pp. 655-667.

Schroeder, R. (2010). *Being There Together: Social interaction in shared virtual environments*. New York: Oxford University Press.

Short, J., Williams, E. & Christie, B. (1976). *The social psychology of telecommunications*. London: Wiley.

Stebbins, R. A. (2001) *Exploratory research in the social sciences*. Thousand Oaks, California: Sage.

Steinkuehler, C. A. & Williams, D. (2006). Where everybody knows your (screen) name: Online games as "third places". *Journal of Computer-Mediated Communication* 11(4), 885–909.

Street, G. (2011). Interview in MMO Champion. Accessed on 20th December 2011 from http://www.mmo-champion.com/content/2588-Interview-with-Greg-Street-%28Ghostcrawler%29-Winter-Veil-EU-Blizzard-Sale-Fan-Art .

Taylor, T. L. (2006). *Play between worlds: exploring online game culture*. Cambridge: MIT Press.

Terrano, M. (2003). Lessons from Life: Designing more immersive games. In *Proceedings of the Australian Game Developers Conference*, Melbourne.

Toups, Z. O., Kerne, A., & Hamilton, W. A. (2011). The Team Coordination Game: Zero-fidelity simulation abstracted from fire emergency response practice. *ACM Transactions on Computer-Human Interaction*, 18(4), 23.

Turner, J.W. & Reinsch, N.L. (2007). The business communicator as presence allocator: multicommunicating, equivocality, and status at work. *Journal of Business Communication* 44(1). 36-58.

Wadley, G. R. (2011). *Voice in virtual worlds*. PhD thesis, Department of Information Systems, The University of Melbourne.

Wadley G., Gibbs M.R., Hew, K. and Graham C. (2003) Computer Supported Cooperative Play, "Third Places" and Online Videogames. In *Proceedings of OZCHI 2003*, University of Queensland, Brisbane, 26-28 November, pp 238-241.

Wadley, G. R., Gibbs, M.R. and Hew, K. (2005). Factors Influencing Users' Decisions to Adopt Voice Communication in Online Console Games. *International Journal of Advanced Media and Communication* 1 (1), 41-58

Wadley, G., Gibbs, M.R. and Benda, P. (2005). Towards a Framework for Designing Speech-Based Player Interaction in Multiplayer Online Games. In *Proceedings of*

IE2005: The Australasian Conference on Interactive Entertainment Sydney (23-25 Nov), 223-226.

Wadley, G. R., Gibbs M.R. and Benda P. (2007) Speaking in Character: Using Voice-over-IP to Communicate within MMORPGs. *Proceedings of the Fourth Australasian Conference on Interactive Entertainment* (IE2007). Melbourne, Australia: School of Creative Media, RMIT University, 8 pgs.

Wadley, G., Gibbs, M. R., & Ducheneaut, N. (2009). You can be too rich: mediated communication in a virtual world. In *Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group* (pp. 49-56). ACM.

Wadley, G. R. and Ducheneaut, N. (2009). The 'out-of-avatar experience': object-focused collaboration in Second Life. In *Proceedings of the European Conference on Computer-Supported Cooperative Work* (ECSCW 09), Vienna

Wadley, G. R. and Gibbs M. R.(2010). Speaking in Character: Voice Communication in Virtual Worlds in William Sims Bainbridge (ed.) *Online Worlds: Convergence of the Real and the Virtual*, Springer: London.

Walther, J. B. (1996). Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication Research*, 23(1), 3–43.

Williams, D., Ducheneaut, N., Xiong, L., Zhang, Y., Yee, N. & Nickell, E. (2006). From tree house to barracks: the social life of guilds in World of Warcraft. *Games and Culture* 1(4) 338-361.

Williams, D., Caplan, S. & Xiong, L. (2007). Can you hear me now? The impact of voice in an online gaming community. *Human Communication Research* 33(4). 427-449.

Woodruff, A. & Aoki, P. M. (2003). How push-to-talk makes talk less pushy. In *Proceedings of the 2003 international ACM SIGGROUP conference on Supporting group work* (pp. 170-179). ACM.

Xu, Y., Cao, X., Sellen, A., Herbrich, R. & Graepel, T. (2011). Sociable Killers: Understanding Social Relationships in an Online First-Person Shooter Game. In *Proceedings of the 2011 Conference on Computer Supported Collaborative Work (CSCW'11)*, New York: ACM Press, pp. 197-206.

Yee, N. (2006). The demographics, motivations and derived experiences of users of massively-multiuser online graphical environments. *Presence: Teleoperators and Virtual Environments* 15(3). 309-329.