Mundane metaverses: communicating by voice in virtual worlds

Greg Wadley
Interaction Design Group
The University of Melbourne
Parkville, Australia
(613) 8344 1586
greg.wadley@unimelb.edu.au

ABSTRACT

Users of online virtual worlds such as MUDs and MMORPGs have traditionally communicated by typing text messages. Recently voice-over-IP (VoIP) has become a popular alternative to text communication. We are interested in how VoIP changes the experience of virtual worlds. This paper discusses research into how people are using VoIP in online worlds. It is argued that some of the problems of voice communication arise because VoIP channels disrupt the boundary between a virtual world and its users' real-world settings.

Categories and Subject Descriptors

H.5.3 [**Information Interfaces and Presentation**]: Group and Organization Interfaces – *Synchronous interaction*

General Terms

Human Factors

Keywords

Virtual worlds, voice-over-IP, VoIP, CMC, MMORPGs

1. INTRODUCTION

Since 2003 we have been studying the use of voice-over-IP (VoIP) technologies for user-to-user communication in online virtual worlds. Users of these recreational systems have traditionally communicated with each other by typing messages, in a variety of synchronous and asynchronous formats. But recently it has become possible to communicate in online worlds by speaking, and this has become a popular alternative to typed text. For example, many players of the massively multiplayer online role-playing game (MMORPG) *World of Warcraft* have adopted third party VoIP products for intra-team coordination and socializing. The vendors of the virtual world *Second Life* added a VoIP feature during 2007, sparking controversy among users. My research question is: how does the adoption of VoIP as a replacement or supplement to typed text affect the experience of virtual worlds ?

Text-based computer-mediated communication (CMC) and voice-based telecommunication are routine technologies for many people. But in 3d graphical game worlds they find a new and in some ways unique use context. Virtual worlds are nearly always used for recreation. "Tasks" are defined by game rules and user consensus, yet are taken very seriously by many users. Communication channels need to be shareable by dynamically changing groups that range in size up to 60 or more. Yet while users are engaged in a fictional world, use takes place in real world settings which can disrupt and be disrupted by the users' actions.

A large body of research dating back to the mid 20th century asks why people choose a particular communication medium, and what effect this choice has on their subsequent social interaction. It is uncertain how applicable those findings are to a virtual world setting. Our research at the IDG aims to bridge this gap by using ethnographically informed methods to find out how virtual worlds users are using VoIP, the benefits they enjoy and the problems they encounter.

2. VIRTUAL WORLDS

Online virtual worlds have existed since the "multi user dungeons" (MUDs) of the 1980s. Those systems allowed users to connect to game servers via a local terminal or dialup connection and engage with other users in an imaginary world. Engagement both with the virtual geography and with fellow users was text-based. There were no screen graphics and the world was presented via description: "You walk through the door into a room. There is a table to your left …"

The first popular multi-user worlds represented as 3d graphics were the networked computer games of the mid 90s such as *Doom* and *Quake*. Users of these "first-person shooter" (FPS) games connected their PCs over local area networks, nominating one PC as the server, and played together in the server's virtual world as team-mates or opponents. Users were usually co-located in the same "real world" room and did not need telecommunication or CMC: they simply spoke to each other. Later these games were Internet-enabled, allowing geographically-dispersed users to play together, and mandating CMC. Users communicated by typing text messages which were displayed on other users' screens by overlaying the text onto the graphical display.

The developers of massively multiplayer online role playing games (MMORPGs) applied the technology of Internet-enabled FPS games to the older MUD concept. The first popular graphical MMORPG, Sony's *Everquest*, brought together many thousands of users to explore, play and communicate in the same virtual world. The most popular MMORPG, *World of Warcraft* (WoW), introduced in 2004, boasts eight million users playing in the same virtual world and is considered by many to represent the mainstreaming of this genre. WoW users typically band together in guilds to cooperate on difficult quests and to socialize. Guild sizes vary up to several hundred users, though 30 is typical. Guilds have complex internal dynamics and are often hierarchically organized. Many users nominate social interaction as a compelling feature of the game. About a third of existing WoW guilds represent "off-line" friendship groups; the rest are users who have met in-game [18].

Virtual worlds such as *There, Active Worlds* and *Second Life* (SL) lack competitive play of the kind traditionally associated with videogames. In these worlds the focus is on role-play and social interaction, and in SL, building and trading in-world objects. Users are represented as graphical avatars with pseudonyms. SL has offered synchronous text chat between users since its introduction in 2003, and a VoIP channel was added during 2007. User interaction in SL is mostly

synchronous, with a focus on events, meetings, real-time discussions, online teaching, and performance. There are currently nine million registered users of SL, though most of those accounts are inactive and the number of regular users is much smaller. SL received significant mass-media attention during 2007 which focused on users' making money by selling virtual land and objects to each other. Researchers are attracted to virtual worlds for the novelty of this form of online interaction and for the ability of virtual worlds to act as experimental labs for social science and economics [4].

3. VOIP IN VIRTUAL WORLDS

Following widespread uptake of broadband Internet, FPS players adopted third-party VoIP products such as *TeamSpeak* and *Ventrillo* in order to speak in real time with team-mates in online games. These voice channels are configured like two-way radios or audio-conferences: usually all members of a team are connected to the same channel and when any user speaks, all team-mates hear them. There seems to be a natural fit between "virtual two-way radios" and games that emulate real-world activities, such as combat, in which radio communication is useful [15]. Players of fast-paced team games have learned that communicating by voice in real time gives them an advantage over teams typing text messages, and this has led to fast adoption of VoIP. Communicating by speech frees up hands for controlling character movement and fighting and allows more efficient coordination of tactics [16].

Game developers responded to the popularity of third-party products by adding VoIP functionality to FPS games such as *CounterStrike Source*. Microsoft's *Xbox Live* game network, released in 2003, offered a voice channel but no text channel. (Text messaging was added to Xbox Live in 2007 using a USB keyboard that attaches to the game console.)

This pattern of adoption has been replayed in the MMORPG genre. In particular, the younger WoW demographic, many of whom also play FPS games, have appropriated the VoIP products they were already using in the FPSs. VoIP has become so popular that many WoW guilds now insist that members use it. Developers have responded by incorporating VoIP into the next generation of MMORPGs such as *Dungeons and Dragons Online* (DDO), released in 2006. In 2007, Blizzard announced that they will add VoIP features to WoW, obviating the need for third-party products.

The eagerness with which VoIP has been adopted for online games suggests that speaking is clearly superior to typing for user-to-user communication. Yet research into use of the *Xbox Live* voice channel [16] and networked PC games [7] found that the usability and sociability of voice was situation-dependent. Shared voice works best when used by a small group of players who already know each other prior to use, and are playing a fast-paced team game in which players must coordinate tactics efficiently. When too many people use one voice channel it is prone to congestion and confusion (though [6] describes one solution to this). The ease of transmitting sound makes VoIP channels easy to abuse. Some people don't like to use voice with strangers. Users often find it difficult to connect the voices in their headphones to the avatars on their screen, and this is worse when they don't know the other players. Voice transmissions, unlike text, can be received not just by those playing the game but by people co-located with players (in the real world) such as family members and co-workers. Conversely, sound from the player's surroundings can be transmitted into the game, leading to unintended breaches of privacy and other problems. Users sit at the boundary between their virtual world and a real world setting, and a voice channel

can violate this boundary. The following diagram shows two users communicating by voice within a game world, while the people co-located with each player are effectively part of the channel as well.
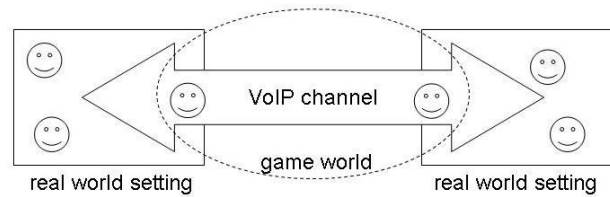
Figure 1: A VoIP channel connects two real-world settings

MMORPGs differ from FPSs in ways that may affect the usefulness of voice communication. While team "raids" are similar to the virtual combat of an FPS, MMORPG players also spend time socializing and trading with team mates and passers by. The social structures and interaction in MMORPG worlds are more complex than in earlier games. Because MMORPG worlds are persistent, asynchronous as well as synchronous communication is required. Users form short-term associations to perform game tasks may include strangers. Users value the ability to communicate not only with team-mates but with any player they might encounter on their travels in the game. Older MMORPGs supported these different communication tasks by providing multiple text channels (one-to-one, vicinity, party, guild etc). It is not clear how best to handle this in a VoIP environment.

While many have embraced VoIP, some virtual world users and commentators are concerned that it will affect privacy and pseudonymity. A voice message conveys more information about the person transmitting it than does a text message. While this property of voice underlies some of its advantages, it is potentially a weakness as well. One prominent author has suggested that transmitting the natural voices of users within an MMORPG will detract from users' ability to play characters, and that text should remain the preferred medium in role-playing environments until technology for changing the sound of voices is better developed [1]. By making a player's gender, age and ethnicity clear, real-world prejudice, harassment, domineering and hierarchies which were previously minimized by the pseudonymity of virtual worlds – and which some virtual world users seek to escape - may consolidate within virtual worlds. Identity exploration of the kind described in [13] may become difficult. People who embrace virtual worlds because a disability or abnormality stops them fully engaging in the real world may find this avenue for social participation cut off: the speech-impaired are an example. Although conversing using voice may increase trust among people who do it [2], it may also create mistrust of people who, for reasons such as the above, choose not to use voice.

Online worlds represent a unique combination of fictional identities and narratives played out in a constructed world, by real people with real motivations located in real-world settings. The anonymity and social distance that for many online communicators is a problem to be solved is used as a resource in virtual worlds to enable role-play. User-to-user communication lies right at this intersection of reality and fantasy. The introduction of VoIP could change virtual worlds dramatically. We are in the right place at the right time to observe these changes and receive the social, psychological and technological insights they offer.

4. PRIOR RESEARCH

A large body of research exists on why people choose particular communication media and the consequences of those choices for subsequent social interaction. But care needs to be taken when applying CMC and telecommunications research to a virtual world setting. Many CMC studies have addressed workplace scenarios, yet work-related concepts such as task, cost, efficiency and effectiveness do not translate straightforwardly to recreational worlds. Users of virtual worlds are engaged in a fiction, yet their actions drive the narrative. People are attracted to the social opportunities offered by virtual worlds, yet often interact with users they will never meet in real life. Human users pilot avatars whose appearances are tailored by them and may represent aspects both of the users and the characters they play [19]. Few other communication systems have featured avatars.

Telecommunications researchers have compared shared-video and shared-audio channels to face-to-face meetings [12]. More recently researchers have studied the use of text-based CMC systems. Media richness theory arrays media along a spectrum from socially-rich face-to-face to leaner written communication, arguing that richer media communicate more social presence and should be chosen when available [5]. Social-influence and critical-mass [9] theories suggest that choice of medium may be determined not only by the medium's properties but by a user's organization, employer or peers. Any of these forces might cause virtual world users to adopt VoIP. Researchers have examined the use of low-social-presence media for deception [3], which may be relevant here if role-play can be considered a kind of deception. CMC researchers have asked whether socially anxious people may prefer text to voice in computer-mediated [8] or mobile phone communication [11], with ambiguous results.

Few researchers have studied VoIP in a virtual world context. A recent study within WoW used a questionnaire to find that team-mates who communicated by voice and text became happier, and liked and trusted each other more, than those who communicated only by text [17]. However it is uncertain how subjects used voice in different gameplay contexts, how often they reverted to text, whether they experienced any problems with voice, how it affected their success in the game, or whether different player types had different media preferences. Nor did the study test the efficacy of voice for large groups or for users who did not already know each other.

5. STUDYING VOIP USE WITHIN VIRTUAL WORLDS

In 2003 we studied user reactions to the voice channel in the Xbox Live game network [16]. We hypothesized that the rich communication environment provided by VoIP and 3d graphics would encourage users to regard online virtual worlds as "third places" in the sense of [10]: settings outside the home and workplace where sociable conversation, a playful mood, and celebration of individual personality are key.

We found that Xbox Live users wanted online community, and to know something of the identity of the people they were conversing with. They preferred playing with people to playing against computer-generated characters: "It's the social interaction you want." We observed players who were fully engaged in conversation to the point of ceasing to play their game. However, some encountered other players online who dominated the voice channel with incessant chat, verbal abuse, or loud music. Other had problems identifying who was speaking and where speakers were in the game world.

In 2005 we studied the use of the *SpatialVoice* system by a team of users playing a regular weekly online FPS game [6]. Whereas most game VoIP systems emulate a two way radio, *SpatialVoice* emulates the transmission of sound in air: when players position their avatars close together in the game they can hear each other's voices, while clarity and volume attenuate as the avatars move apart. Although *SpatialVoice* represented a restriction of the usual "radio" configuration, it helped reduce channel clutter when many people were using voice, and acted as a "relevance filter", because users received communication about local action only and knew that any voice they heard must belong to a player who was nearby in the game.

In 2006 we observed use of VoIP in two different MMORPGs over a period of two months [14]. Participants played in their own homes under their normal playing conditions. They kept diaries and participated in interviews and focus groups. We asked whether participants preferred voice or text, whether one medium was better suited to particular types of gameplay, whether they had encountered episodes in which voice was especially useful or problematic, and whether there were aspects of the VoIP interface they would like changed.

We found that the different abilities of VoIP and text to convey social presence meant that either medium might be more appropriate depending on context. The suitability of each medium depended on a dynamic compromise between the need for users to know and understand each other and their desire for privacy and identity-play. We concluded that text may persist as the best medium in some MMORPG scenarios.

MMORPG voice channels were sensitive to different users transmitting at unequal volumes (for which there is no analogy in text communication). Network lag made voices choppy and led users to talk over each other, forcing them to put more effort into channel coordination, which reduced the ease of voice communication. "I find there's a lot of stuttering. It's not really lag. You'll start to say something, but you'll stop because someone else is talking, because they didn't realize [you were talking], because of that time delay."

Other unexpected problems arose. MMORPG players using voice were confronted with the problem of how to pronounce character names and other words drawn from imaginary languages used in these games. One participant was uncertain at first how to pronounce game jargon: "Some things you type all the time, but you never actually say out loud. The first time I go to say it, I stumble over it. Do I say 'exp' or 'X. P.' or … ".

Communicating by voice made participants more sensitive to knowing who was able to hear them. Some wanted to be able to speak to strangers in the game world, while others preferred to speak only with people they knew, or felt a voice channel that allowed anyone to talk to anyone would be abused or dominated by particular individuals.

Some participants bemoaned the inability of VoIP to support asynchronous communication or allow users to "scroll back" through a conversation, though some thought an "answering machine" or "time shifting" facility would overcome this.

Some felt that the emotional impact of voice made it more prone to flame wars. However others felt that users were inclined to be more polite when speaking than when typing: "There's certainly something a bit different about shouting 'shut up you d---head' into the microphone and typing something like that." Some thought that the immediacy of speaking meant that swearing was likely to happen before a speaker thought about the consequences. Users couldn't be

certain who might hear an utterance: for example a team-mate's children might be listening.

Our participants described incidents in which sounds such as breathing, eating, household noise, speech from family members, TV, and music were accidentally transmitted via microphones into the game's voice channel. People playing with friends they knew well could recognize whose spouse or children were making noise. One reported an incident in which he inadvertently placed his headset beside his baby while comforting her, thereby broadcasting the baby's crying to his team-mates. One participant developed a cough that made VoIP temporarily unusable for him. Another reported overhearing young players being told by their mothers to get off the computer, followed by the player announcing "gotta go now" into the voice channel.

Conversely, speech intended for transmission into a virtual world can interrupt a speaker's own household. Participants reported playing differently depending on whether family members were at home: in particular if small children were around they used voice differently. One felt uncomfortable speaking if his family were present, but only in situations when he didn't know the players he was speaking to: "I think it's important that text is an option too, for when somebody else is home and you just can't talk."

6. CONCLUSION: MUNDANE METAVERSES

Virtual worlds such as *Second Life* are similar to online games in that users communicate within a 3d graphical world, but different in that there is little need for fast-paced coordination, the requirement that motivates much VoIP use. Pseudonymity and role-play are cherished by many SL users. During 2007 a VoIP system was added to SL: its impact will be interesting to observe. We plan to study this channel during 2008. Interviews and focus groups can be carried out within SL itself, making the communication channel itself a research tool. The recent media interest in SL suggests that, like MMORPGs, use of this virtual world will become a routine activity and communication medium for many people.

Users of online virtual worlds sit at the intersection of two locales: one virtual and one real. Online worlds are often compared to the "metaverses" of science fiction, yet users experience them routinely in domestic settings. Some of the problems of voice communication are caused by its rupturing of the boundary between the virtual world and its users' surroundings.

7. REFERENCES

1. Bartle, R. Not yet you fools! Game + Girl = Advance, 2003.

2. Bos, N., Olson, J., Gergle, D., Olson, G. and Wright, Z. Effects of four computer-mediated communications channels on trust development Proceedings of the SIGCHI conference on human factors in computing systems, Minneapolis, Minnesota, USA., 2002.

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

4. Castranova, E. The Right to Play. New York Law School Law Review.

5. Daft, R.L. and Lengel, R.H. Organizational information requirements, media richness and structural design. Management Science, 32 (5). 554-571.

6. Gibbs, M., Benda, P. and Wadley, G. Proximity-based chat in a first person shooter: Using a novel voice communication system for online play Proceedings of the 3rd Australasian Conference on Interactive Entertainment, Perth, 2006.

7. Halloran, J., Fitzpatrick, G., Rogers, Y. and Marshall, P. Does it matter if you don't know who's talking? Multiplayer gaming with voiceover IP In Proceedings of CHI 2004, ACM Press., 2004.

8. Madell, D. and Steven Muncer, P.D. Rapid Communication Internet Communication: An Activity that Appeals to Shy and Socially Phobic People? Cyberpsychology and Behaviour, 9 (5).

9. Markus, M.L. Toward a "critical mass" theory of interactive media. in Fulk, J. and Steinfield, C.W. eds. Organizations and Communication Technology, Sage Publications, Newbury Park, 1990, 194-218.

10. Oldenburg, R. The great good place : cafes, coffee shops, community centers, beauty parlors, general stores, bars, hangouts, and how they get you through the day. Paragon House, New York, 1989.

11. Reid, D.J. and Reid, F.J.M. Text or Talk? Social Anxiety, Loneliness, and Divergent Preferences for Cell Phone Use. Cyberpsychology and Behaviour, Volume 10, Number 3, 2007.

12. Short, J., Williams, E. and Christie, B. The social psychology of telecommunications. Wiley, London, 1976.

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

14. Wadley, G., Gibbs, M. and Benda, P. Speaking in character: using voice-over-IP to communicate within MMORPGs Proceedings of the Fourth Australasian Conference on Interactive Entertainment, Melbourne, 2007.

15. Wadley, G., Gibbs, M. and Benda, P. Towards a framework for designing speech-based player interaction in multiplayer online games Proceedings of the 2nd Australasian Conference on Interactive Entertainment, Sydney, 2005.

16. Wadley, G., Gibbs, M., Hew, K. and Graham, C. Computer supported cooperative play, "third places" and online videogames Proceedings of the Australasian Computer-Human Interaction Conference, 2003.

17. Williams, D., Caplan, S. and Xiong, L. Can you hear me now? The impact of voice in an online gaming community. Human Communication Research, 33 (4). 397-535.

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

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