

Speaking in character: Using voice-over-IP to communicate within MMORPGs

Greg Wadley, Martin Gibbs, Peter Benda

Interaction Design Group

The University of Melbourne

111 Barry St Carlton 3053

[greg.wadley, martin.gibbs, pbenda]@unimelb.edu.au

ABSTRACT

While voice-over-IP has long been favoured as a communication medium by players of team-based online shooter games, it has recently also been appropriated by players of MMORPGs, and some recent MMORPGs have included voice facilities in the game software. However voice communication has provoked controversy among players and designers, some of whom believe that it is not suited to some of the communication tasks required in this genre of games, such as role-play, coordination of large groups, and interaction with strangers. Little research has been published on VoIP use in MMORPGs. We studied the use of voice by three groups playing *Dungeons and Dragons Online* and *World of Warcraft* over a period of three months. The players kept diaries, were interviewed individually, and participated in focus groups. We organized this data into themes which are presented here. We discuss our findings with regard to prior research into computer-mediated communication..

Categories and Subject Descriptors

H.5.3 [Information interfaces and presentation]: Group and Organization Interfaces – *collaborative computing, computer supported collaborative work, synchronous interaction*

General Terms

Design, Human Factors

Keywords

MMORPG, Voice-over-IP, VoIP, computer mediated communication, CMC, media effects, *Dungeons and Dragons Online*, *World of Warcraft*

1. INTRODUCTION

The first generation of networked first-person shooter (FPS) games included facilities for players to communicate with each other by typing text messages. Following widespread uptake of broadband Internet, FPS players began to adopt VoIP products such as *TeamSpeak* and *Ventrillo* in order to speak in real time with team-mates while playing. 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 member speaks, all others hear them. There seems to be a natural fit between these “virtual two-way radios” and games that emulate real-world activities (such as infantry combat) in which two-way radios are useful [22]. Communicating by voice in real time gave players of fast-paced team games an advantage over teams typing text messages, motivating widespread adoption of VoIP by FPS players. Game developers responded to the popularity of these products by adding VoIP functionality to subsequent FPSs such as *CounterStrike Source*. The *Xbox Live* console networking

system, released in 2003, provided a voice channel but no text channel¹. It would seem that gamers have voted for voice as their preferred method of communication in online FPS games.

A similar pattern of adoption appears to be occurring within the massively-multiplayer online role playing game (MMORPG) genre. Early graphical MMORPGs such as *Everquest* did not have facilities for voice communication. Many MMORPG players began to appropriate the VoIP products already popular in FPS games. Developers have subsequently incorporated VoIP into the next generation of MMORPGs such as *Dungeons and Dragons Online* (DDO), *Lord of the Rings Online* (LOTRO) and *EVE Online*. Recently, Blizzard announced that they will add voice features to *World of Warcraft* (WoW), the most popular game in this genre.

The eagerness with which VoIP has been adopted by users suggests that shared voice channels are a superior way to communicate in online games. Yet research into use of *Xbox Live* [14] and networked PC games [12] found that the usability and sociability of VoIP was situation-dependent. VoIP works best when used by a small group of players who knew each other prior to use and were playing a fast-paced game in which they needed to coordinate their tactics efficiently. If any of these criteria are not met, the usefulness of voice may be reduced. When too many people use the one voice channel it is prone to congestion and confusion (though [11] describes one solution to this). The ease of transmitting makes the channel easy to abuse. Some players prefer to communicate by text when they don't personally know the people with whom they are playing [14]. Players 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 and are unable to recognize their voices [12]. In games which are multiplayer but not team-based, such as racing sims, voice seems to have little value other than for trash-talking, which is not desired by all players. Speech, unlike text, can be received not just by those playing the game but by people co-located (in the real world) with players, such as family members or co-workers. Conversely, sound from the player's surroundings, such as household noise, may be transmitted into the game, leading on occasions to unintended breaches of privacy and other problems [14].

MMORPGs differ from FPSs in ways that may affect the usefulness of voice. There is a greater variety of game activity. While team raids are similar to the “virtual combat” of an FPS, players also spend time exploring the game world, socializing with team mates or passers-by, and depending on the game, building structures and trading possessions. While in both genres players need to cooperate and communicate to achieve

¹ Text messaging was added to *Xbox Live* in 2007, using a USB keyboard that attaches to the *Xbox 360* console.

goals, the social interaction in MMORPG worlds populated with millions of players over an extended period of time is more complex. For many MMORPG players the game is primarily a social experience, making convivial communication particularly important [23]. Most MMORPG players join guilds which are often larger than typical FPS teams and have hierarchical structure [27]. Yet at a given time usually only a subset of a guild is logged in, so that asynchronous communication is required. Players also form short-term associations which may be drawn from their guild ("raiding parties") but may include strangers ("pickup groups"). (In this paper when it is unnecessary to distinguish between guilds, raiding parties and pickup groups, we use the word "team".) Finally, MMORPG players value the ability to communicate not only with team-mates but with any player they might encounter on their travels in the game world.

Communication therefore occurs on both long and short time scales, over long and short game-world distances, synchronously and asynchronously, and between players both known and unknown to each other. Older MMORPGs have supported these different communication tasks by providing several different text channels (whisper, vicinity, party, guild etc). It is not clear how best to approach this in a VoIP environment.

The introduction of voice communication into MMORPGs has provoked controversy. While many have embraced voice, others are concerned that it will reduce the privacy and pseudonymity of user participation. 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 users' natural voices within an MMORPG will detract from their ability to play a character, and that text should remain the preferred medium until technology for changing the sound of voices is better developed [2]. Designers, academics, players and commentators have debated on the TerraNova and Slashdot blogs the possible drawbacks of implementing voice in virtual worlds. 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 MMORPG players seek to escape – may consolidate within virtual worlds. The separation between human players and the in-world characters they play is narrowed, and this may reduce the quality or even the possibility of role-play. Identity exploration of the kind described in [1] and [21] 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: hearing- and speech-impaired people are two examples. Although conversing by voice may increase trust among people who do it [3], it may also create mistrust of people who, for reasons such as the above, choose not to use voice.

A panel at the Spring 2006 VON conference (reported at http://www.christine.net/2006/03/the_impact_of_v.html) listed the major problems associated with voice in MMORPGs as being: synchronicity, identity and privacy, disruption of people co-located with the player, voice quality, the need for speakers to already know each other, congestion due to many users, difficulty localizing a speaker, difficulty socializing with passers-by, the "military two way radio" metaphor, and mismatch of players' voices to their characters (list paraphrased by current authors).

2. PRIOR RESEARCH ON MEDIA USE

Voice is a recent arrival in MMORPGs and little research has been published on how it is being used. In a recent quantitative study, researchers provided VoIP to existing small groups in WoW for one month and measured its effect on participants' social attitudes and relationships using a questionnaire [26]. They found that team-mates who communicated by voice and text during the trial liked and trusted each other more, and became happier and less lonely, than those who communicated by text only. However it is uncertain how the players used voice in different gameplay contexts, how often they reverted to text, whether they ever experienced problems with voice, how it affected their success in the game, or whether different player types had different media preferences. This study did not test the efficacy of voice for large groups or for players who did not already know each other.

The broader question of why people choose a particular communication medium, and the consequences of that choice for the quality of their subsequent interactions, has been addressed by researchers of "media effects" on human communication. Short et al's seminal review of telecommunication research [19] compared, in a workplace context, shared-audio (voice) and shared-video channels to face-to-face meetings, with respect to criteria such as efficiency, persuasiveness, conflict resolution and group cohesion. The authors concluded that these media differed in how well they conveyed the "social presence" of communication partners. In subsequent studies of computer-mediated communication (CMC), Daft and Lengel [6] extended this reasoning to arrange media in decreasing order of richness (ability to convey social presence): face-to-face, telephone, personal written document (e.g. email), impersonal document (e.g. memo), and numerical document. The authors argued that people should choose the richest medium available, especially in situations of ambiguity. However both the social-presence and media-richness theories acknowledged that leaner media may be more efficient for communication that emphasizes transmission of information, during which social presence might be a distraction.

Data contradicting the prediction that people rationally choose rich media led subsequent researchers to criticize media richness theory and propose alternative prescriptive and descriptive models of media use. The social-influence [10] and critical-mass [17] theories proposed that choice of medium is determined less by the medium's properties than by choices already made by a person's organization or fellow communicators. Other research, e.g. [29], has argued that medium suitability is largely task-dependent. Media synchronicity theory analyses media along multiple dimensions: their ability to support feedback, symbol variety, parallelism, rehearsability and reprocessability [7]. For example, while email is limited in its support for symbol variety and feedback, it supports large numbers of simultaneous users, and allows them to rehearse (edit) a message before transmitting, and to store and re-read messages. Lober et al conducted a series of experiments comparing voice-over-IP and instant messaging in different settings. They found that text scaled better with group size [16] and that text and voice were equally efficient for distributing information between group members, but that voice was faster than text for resolving ambiguity [15].

Carlson and Zund [5] argued that the social richness of a medium depends on the users' prior experience with the medium and each other. Walther [24] argued that people may be motivated to work around the shortcomings of lean media to create rich communication. Ngwenyama and Lee [18] argued

that the richness of a message is produced by the social context of the conversants. They found that people engaged in CMC constantly interpret not only the meaning but the validity of the messages they receive, with the identity of the sender in mind. Erickson and Kellogg [8] introduced the concept of social translucence to argue that a medium should transmit enough – though not all – information about the actions and status of its users.

Other researchers argued for a shift in focus from users' choice of individual media to their use of a collection of media [25], and found that people linked by strong ties use more media to communicate [13]. Some argued that people wishing to engage in deception, or "impression management" [4], or who are shy [20], might choose a medium with low social presence such as text, and that while text-based communication "flattens" hierarchies, audio conferences exaggerate them so that high-status individuals to dominate shared audio channels [9].

Care needs to be taken in applying existing CMC research to a MMORPG scenario. Whereas most CMC studies have taken place in (real or simulated) workplace settings, MMORPGs are a recreational technology. Work-related concepts such as task, cost, efficiency and effectiveness do not translate perfectly to games. MMORPG players are thrown into online interaction with people they may never meet in real life. MMORPG groups are typically larger than those used in CMC research, and neither shared-video channels nor face-to-face meetings are usually available to MMORPG users. However the graphical representation of players as avatars in the game-world shares some of the characteristics of video channels. While avatars lack most of the visual characteristics of the person controlling the avatar, recent research indicates that some "real world" interpersonal phenomena such as gaze direction are preserved in-game [28].

In the study reported here we used ethnographically-informed methods to observe players using voice in MMORPGs over an extended period, in order to develop a richer understanding of how VoIP is used in MMORPGs and allow for future theory development.

3. OUR STUDY

We arranged for three groups of five participants to use VoIP in an MMORPG over a period of two months. All participants were experienced MMORPG players, and many had already used VoIP in online FPSs. Two of the participant groups played *Dungeons and Dragons Online*, the first MMORPG to include VoIP facilities within the game software. 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 DDO before. A third group were already regular users of third-party VoIP products in *World of Warcraft*, *Everquest* and other MMORPGs: they continued to do this during the study and contributed to our data collection.

Participants played in their own homes during the trial, 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 Results section below.

The DDO voice channel is configured such that when individual players joined a pickup group they could speak with team-mates in a "two way radio" format. Players couldn't speak with non-team-mates, or with ex-team-mates after the team had disbanded. DDO didn't distinguish between pickup groups and guilds, though it supported long-term associations with a friends-list feature.

4. RESULTS

4.1 First encounters with voice

Most participants had trouble getting voice to work at first, but were satisfied eventually. One participant never got it to work: he could hear others, but no-one could hear him. Some participants found the user interface of the DDO voice channel hard to understand. One said that the overall quality was satisfactory though minor problems occurred constantly.

Participants reported that many of the DDO players they encountered early in the study appeared either not to have voice hardware, or to be unable to use it, or to have it switched off or incorrectly configured. Some reported spending their first couple of play sessions frequently saying "Hello, hello?", unsure whether other players could hear them.

One participant found voice intimidating at first. Another reported that he became more comfortable and familiar with voice during the period of the study. Another felt that problems with use of voice would be ironed out by the player community over time.

Participants reported that at the start of a play session, team members needed a coordination phase during which the status of each player's voice setup was fixed, using spoken or typed feedback from team-mates. This process was more arduous in pickup groups, where players were unfamiliar with each other and the quirks of each other's voice hardware. Participants reported gameplay sessions in which they had assumed they could or couldn't be heard, only to find out later that their assumption was wrong. Often, only some of the members of a team had voice working while the rest used text: this led to fragmentation of the party into sub-teams based on medium.

A voice channel is more sensitive to network lag than is a text channel. A transmission delay a fraction of a second long can make a player's voice choppy and hard to understand. A longer lag can cause people to unintentionally talk over each other or lose track of a conversation. One participant reported a lag of two minutes during one game session, with comical results: his teammates kept him in the know by telephone. As one participant reported, lag could lead to a breakdown of communication norms:

“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. You think, I’ll be polite and not interrupt. Then at that stage they hear your voice and they stop. So you get these jerky teams where people think, oh screw it, I’ll just talk over the top of everyone.”

The usability of a voice channel is sensitive to the relative volumes at which different players hear each other. There is no direct analogy to volume in a text channel. Many variables control the loudness with which a given utterance is heard by others. Some of these cannot be controlled or even measured by the game software. Some participants commented that they wanted to normalize the volumes of individual team-mates; however DDO did not provide a way to do this.

The “chattiness” of the other players encountered by our participants varied. Some participants joined pickup groups in which some players would rarely speak, though they were apparently listening to the voice channel and might comment on voice utterances using text.

Some participants struggled with how to pronounce DDO character names, which consist of syllables in a fictitious language. One participant heard players shortening or familiarizing unwieldy character names. One participant was uncertain at first how he should 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 ...”

When participants were unable to communicate using the voice channel, they reverted to what one called the “safety net” of the typed-text channel.

4.2 Advantages of voice in MMORPGs

Our participants universally approved of communicating by voice – when the circumstances were right. After technical problems were resolved, participants found speaking to be easier, more natural, and more relaxing than typing. They felt that voice communication enhanced the gaming experience, especially when playing with a good, sociable team. One participant said voice made the game “feel like a living breathing party”. Conversations flowed more freely, 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.

Because speaking freed a player’s hands from typing, participants could communicate while carrying out other game actions such as moving and fighting. Participants reported incidents in which this made the difference between success and failure in achieving game goals.

One participant felt that personalities were broadcast more readily through the voice channel. Participants said that communication for purposes other than achieving game-goals, such as socializing, joking and what the participants called “off topic” chat, was easier with voice and happened more often. Participants felt that if they made a genuine mistake in play and needed to apologize to team-mates, it would be more immediate and believable if spoken.

The voice channel allowed “newbies” to easily ask for help. One participant reported that most of the speaking she

encountered consisted of players helping each other learn the game.

One participant reported an example of voice use that would have been impossible with text. After a group raid, a team member invented a ballad about the raid and sang it to the team. One participant used the voice channel to broadcast a humorous stream-of-consciousness report of her actions. She felt that voice supported “in-context humour” better than text, and enjoyed this. However she noted that “humour works best with people you know”.

There was a fine line between chattiness and over-use. The ease of voice communication made it easy to broadcast too much. A chatty player could easily dominate the channel and annoy team-mates with what one participant described as “verbal diarrhea”.

4.3 Voice is most useful in raids

All participants agreed that voice is better than text during team raids, the part of MMORPG play that most resembles FPS play. “Two-way radio” seems ideally suited to dynamic coordination of a small, mobile, fighting group of players.

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

“We change our tactics on the fly, and coordinate via voice, which would be impossible via text. In WoW, which is text, you have to know your position. Each person had their niche and you had to understand it. If you didn’t fulfill your role you mucked around everybody else. Whereas with DDO you’ve got that real advantage of being able to change tactics on the fly.”

Voice was superior for raid leaders issuing directions to other players, and for wounded players calling 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. However participants noted that a shy player might more easily be pushed into taking a smaller share when negotiation took place using voice.

4.4 Activity outside of team raids

One participant said his favourite use of voice was to chat “off topic” with his friends while they explored the game-world. Some participants felt the best time for off-topic chat was during boring, repetitive play such as trade-skilling and grinding. However the suitability of voice to off-topic communication depended on whether players knew each other well. Participants reported that voice was less effective when players didn’t know each other, and that while most players were happy to speak about game action, only people who knew each other in real life discussed topics unrelated to the game. One participant who usually played in pickup groups reported “occasional off-topic banter at the end of raids”, but “didn’t encounter anything particularly meaningful”. Some players were shy or unwilling to speak in front of strangers. One player likened communicating through text to “putting up a curtain”, in that it obscured personal information and provided more privacy. A small number of participants were not interested in off-topic communication.

Using voice made some participants more sensitive to knowing who was receiving their broadcasts. Some participants

suggested that this would be an important issue if an MMORPG had “vicinity” voice (a “3d” voice channel such as the one discussed in [11]): in this case they would want to be notified when another player’s avatar had moved within hearing range of their conversation.

Participants reported that voice was better for permanent groups than for pickup groups, and that adopting voice had made them reluctant to add strangers to their team. One reported that speech within pickup groups was sparse and instrumental, whereas if a group knew each other “it was like being in a room together and they’d chat about anything”. One participant pointed out that players in a pickup group are meeting three or four people for the first time, all at the same time, and suggested that the game software should help them negotiate this complexity. Participants who mainly played with strangers complained that they often didn’t know who was speaking. DDO provides more help than the third party products in this regard, by lighting an icon beside the speaking player’s name, but by the time participants looked at the list of names the utterance was often finished. To solve this, some participants suggested that a speaking player’s avatar should light up, or a speech bubble appear, or that the sound of a voice be directional.

Some felt that voice would also be useful when speaking to strangers whom they met while traveling in the game world. For example in taverns they wanted to use voice to ask for healing, and then thank obliging strangers, in a friendly, personal way. They felt that “one on one” voice chat with strangers could be implemented like a text “whisper”: a player could click on a passerby’s avatar and proceed to speak to them. One participant felt that teams using voice spent less time socializing with passers-by, and that a one-to-one channel would help solve this.

However some participants felt that an out-of-team voice channel might be abused. They offered as an example the “looking for group” text channel in WoW, which some use to broadcast offensive text. Some participants felt that harassment of female players, racial taunting and other forms of abuse they had witnessed in text channels would be worse in a voice channel which allowed any player to speak to any other player. “You don’t want to go into a game and then get persecuted because you’re a Jew or you’re gay, you’re going there to have fun.” However one female DDO participant said: “Nobody tried to chat me up, nobody made any sexist comments. I had one, but it was a joke.” Some participants expressed annoyance at misuse by younger players of the voice channel. One participant encountered a player who used the voice channel to evangelize religion, and felt this to be an abuse of the channel. She was wary of how a shared voice channel would be used by large groups gathered in taverns or public areas, feeling that someone would probably try to dominate the channel.

4.5 Speaking in large groups

Participants reported that voice was less useful in large groups, because players were more likely to talk over each other and because it was harder to recognize voices. Some DDO participants wanted a “guild chat” mechanism in addition to the provided “raid chat”; though one participant felt this might prove to be “like trying to speak in a crowded room”. One WoW participant reported that his group didn’t try to use voice in groups larger than 6, and felt that waiting for your turn to speak detracted from immersion. Another believed that at most 8-10 people could use the same voice channel simultaneously.

However one participant believed that text channels suffer the same problem when used by large raiding parties, and that voice was still easier overall when coordinating large groups.

One WoW participant reported that of his guild of 70 players, about 40 would usually take part in a raid, communicating using a third-party VoIP product. They minimized channel clutter by restricting it for only leaders and officers to broadcast commands, rather than supporting free-for-all chat.

4.6 Multiple channels

MMORPGs typically offer several different typed-text channels to meet different communication needs, for example guild, raid, one-to-one and vicinity chat. Many players keep several channels open at once and may conduct multiple conversations in different windows simultaneously. Some participants playing WoW reported embarrassing situations where a player typed a message into the wrong channel, such as when team members were discussing a team-mate behind their back. One participant felt that this kind of error would be worse with voice: “Sometimes you’d hear things you didn’t want to hear, or people who were annoying because they were having a long conversation right next to you.”

We asked participants whether they wanted multiple simultaneous voice channels. Some felt it would be impossible to monitor several conversations simultaneously, though it was relatively easy to do in text. Others did want channels, and suggested that the interface be presented as a simple list of screen buttons, one for each channel.

Several DDO participants wanted a guild voice channel, and some wanted all-to-all voice in public spaces such as taverns. “I would like it if I could go to a tavern and hear somebody say, I’m looking for fighters to go into a group. That’s the kind of thing I would find it useful for.” However participants recognized that all-in or vicinity voice chat in public spaces could become a clamour. One suggested that the solution to this was directional sound:

”Voice chat needs to get sophisticated enough so that you’re walking through a market place or something, and you’re like this situation here, where we’re sitting in a room and you can have several conversations going on at once, and you can focus on individual conversations without the other conversation just wiping it out - because of the direction of the voice. ”

Some wanted to be able to send voice “tells” to individuals. Example scenarios were negotiating while trading, and dispersing loot after a raid, during which issues could arise that needed to be discussed without the whole group hearing. It was suggested that tells be implemented so that senders addressed recipients by speaking their name, like the recorded friends-list on some mobile phones. To avoid the possibility of abuse, participants suggested that clicking on an avatar represent a request to chat rather than automatically opening a channel.

One WoW player described his guild’s simultaneous use of several Ventrillo channels: raid, officer, chat, and a channel for each character class. This guild conducted highly coordinated raids with one “general” in overall command, several officers, and a leader for each player class. They considered Ventrillo compulsory for raid members: their attitude was, use voice or don’t come on the raid.

Multiple text channels can be implemented partly because users can keep each channel boxed in a separate window, and scroll

back through message logs to re-read missed messages. Some participants felt it would be impossible to implement a “scrolling” mechanism in voice. Voice could perhaps be recorded, but it would be hard to scan through a conversation. Participants reported a need to scroll frequently, for example when they came in mid-way through a conversation or had to take a break from the game. Participants said they wouldn’t use a recording in this way. Some felt that scrolling and multiple channels were significant advantages for text over voice.

4.7 Asynchronous communication

In a persistent-world game only a subset of a guild may be logged in at a given time, yet activities such as planning and coordinating need to be on-going. One WoW participant also described problems receiving voice chat from guild-mates while in the heat of battle:

“All of these situations where you can’t respond immediately, or it’s too confusing to have people who aren’t there in your party talk to you while you’re in a quest, that’s where text comes into its own. You don’t mind using text for that kind of thing.”

Neither the DDO voice channel nor the third-party VoIP systems used by WoW participants supported asynchronous communication. We asked participants whether this was a problem, and how a voice channel might be designed to support asynchronicity. Participants discussed the possibility of answering machines and voicemail, or being able to “time shift” the voice channel, similar to recording a television broadcast. However participants were apprehensive about checking for and playing back messages. One thought it would be simpler just to ask the sender to send again. “If you knew it had come from Jim, you’d just say ‘Jim, what did you say back there?’”

A further suggestion was that players who missed a message be informed later by a text message on screen. Another was that the game signal a player’s unavailability by responding to voice calls with a canned recording of battle sounds. Some participants suggested that the voice channel could implement a mechanism similar to that in EverQuest, where tells to a player marked away-from-keyboard are bounced.

4.8 Immediacy, emotion, and flame wars

Participants noted that when using text, a player is more likely to think for a moment before pressing send, whereas a player equipped with a microphone is more likely to speak before considering what to say. One participant offered the “Leeroy” clip as an example. Another participant mentioned that recordings of MMORPG players “losing it” at fellow team members were being placed on the Internet for amusement.

Participants felt that the emotional impact of a voice transmission could be greater than that of text, and that this could cause players to get angry, wrecking team dynamics. They felt that this was more of a problem for pickup groups than for people who played together regularly. On the other hand a voice channel could heal rifts as well: a team member who made a mistake could apologize immediately and his sincerity would be perceived.

One participant felt that voice users were more polite than text users. “People treat voice like a phone conversation – they have phone etiquette. There’s certainly something a bit different about shouting ‘shut up you dickhead’ into the microphone and typing something like that.” Though he had seen a lot of swearing in text channels, he heard little on the DDO channel

Participants believed that whether swearing was appropriate depended on whether the users of the channel were close friends. One participant drew an analogy with accidentally bumping into someone on the footpath: one would be wary of foul language if the other person was a stranger in case they took offence.

However some participants agreed that the immediacy of voice meant that swearing was likely to happen before the speaker thought about the consequences. This was dangerous because a speaker couldn’t be certain who might hear an utterance: for example a team-mate’s family might be listening.

4.9 Real world sounds leak into game world

Participants described incidents in which sounds such as breathing, eating, household noise, speech from family members, TV, and music were accidentally transmitted into the voice channel. One participant developed a cough that made voice temporarily unusable for him. Another described a situation in which he temporarily stopped playing to comfort his small child, placing his headset beside the child and thereby broadcasting the child’s crying into the game. Participants reported that when playing with team-mates they knew well they could recognize whose spouse or children were making noise in the background.

One participant reported overhearing younger players being told by their mothers to get off the computer, followed by the player announcing “gotta go now” into the voice channel.

4.10 Game sounds leak into real world

Conversely, sound effects and speech from a game can interrupt a player’s household, and so can the player’s own utterances intended for the game. One participant felt: “I think it’s important that text is an option too, for when somebody else is home and you just can’t talk.” One participant reported adjusting his volume to achieve a compromise whereby team-mates could hear him but he wasn’t disturbing his family.

Two participants reported that they played differently depending on whether family members were at home and able to hear them. In particular, if small children were around they used voice differently. One participant felt uncomfortable speaking if his family were present, but only in situations when he didn’t know the players he was speaking to.

One teenaged participant said that he played in his room, and regarded speaking into a voice channel as being like talking on the phone. He was not concerned about his parents overhearing.

4.11 Does a voice channel support role-play?

Our participants had varying interpretations of what “role playing” meant. One argued that you don’t need to feign an accent or use “Tolkien-speak” in order to role-play, claiming that this was a common misunderstanding. He preferred to define role-play in terms of playing an assigned role in the team. On the other hand most of our participants seemed to define role-play according to whether players were speaking in character - as though they were an elf or dwarf or whatever. Several participants claimed to have encountered few role-players in DDO or WoW. One had encountered a player in a pickup group who “threw in a hearty arrgh now and again”. One believed that all role-players use text:

“I’ve never spoken to anyone who’s actually tried to speak in character, using their voice. It’s rare to see it in

text. I can't imagine anyone trying to pull it off using their voice."

Another thought of role-play in terms of depth of immersion:

"I like to be in character, but not in a heavy-duty way - just pretending it's me that's running around. 'Oh look at that', and 'I just fell in the water again'. But not putting on a voice or trying to be 'ye olde' or anything. I prefer that other people I'm playing with are doing that level of role-play so that when something happens they respond to it as if it had happened, rather than just having a chat."

Most participants however encountered situations where another player's appearance and voice did not match. The most common examples cited were males playing female characters, children playing adult characters, and players with accents. These problems did not appear to be due to prejudice in our participants but due rather to the extra effort required to translate real voices into the game fiction. One player commented: "My character looks scary, but my voice isn't - how's that going to work?" No-one was annoyed by "tough guys with kids voices", though some found it amusing.

Some described their encounters with other players engaged in gender role-play as weird and uncomfortable. Some simply found it confusing. One female participant deliberately created a female character so that her voice would match the character's appearance.

Some of the male participants had played female characters prior to this study, and reported that other players with whom they had grouped for a long time didn't realize they were males until they adopted VoIP in those games. These participants however had also role-played in pen-and-paper games, in which hearing a fellow player's voice was normal, and playing a character of a different gender, race or even species worked perfectly well.

"[in pen-and-paper roleplay] when you play characters you've already got that consent. There'll be no hassles if I roll a female character. They're not going to think I'm queer because I want to play a female character, because you tend to know the people. Whereas when you're online and playing a female character, you don't know how people are going to react. The people we were playing had no issue with it, it was just the surprise that we were actually male. We're destroying the illusion if we start using voice."

One "occasional" role-player used text during this activity, because his collaborators spoke off-topic too often if they used voice. Another was happy to role-play in text but felt it would be embarrassing to attempt it with voice if not everyone in the group was speaking in character.

Most participants had little enthusiasm for the idea of tweaking or affecting their voice. However some said they wanted the ability to sound like the character type they had chosen to play. One participant pointed out that this is not merely a matter of changing tone or accent but also vocabulary, which is probably impossible with current technology.

One participant said that there were some game situations where voice seemed inappropriate; for example engaging in light banter during moments of drama and death.

5. DISCUSSION

Our data confirm that MMORPGs are complex games, played by different people with different motivations in a variety of real-world settings. Data gathered in this and previous studies indicate that there is a paradigm gaming scenario to which VoIP is ideally suited: small groups of people who know and are comfortable speaking with each other, who are simultaneously engaged in the same fast-paced team combat task, whose main purpose for playing is to succeed in the game, and whose physical setting does not cause non-players to be involved in the game-world communication. However this is not the only scenario found in MMORPGs.

Perhaps a third or more of players use MMORPGs as a way to maintain real-world friendship groups [27]. It might be expected that for these people the greater social presence of voice communication would enhance the online play experience. On the other hand, people whose team-mates are not known to them offline might prefer the social distance of a text channel. Groups that play together primarily for instrumental reasons - to successfully engage in raids - might prefer the voice channel's support for faster resolution of ambiguity, as predicted by media richness theory. On the other hand some groups use the voice channel as a one-way broadcast medium when engaged in raids. Voice may be useful in other situations of ambiguity such as trading and negotiating loot.

To an extent, our data support the applicability of existing CMC theory to MMORPGs. In concordance with [16], players reported that the usefulness of voice relative to text decreases as group size increases. Role-play is not dealt with in CMC literature; however one could argue that it is a kind of "impression management", making role-players' preference for text compatible with [4]. Voice channels seem prone to domination by some individuals, as described in [9].

Rather than concluding that either voice or text is the better medium in all situations, we found that the different properties of voice and text make these media more or less suited to different game scenarios and goals. The framework best suited to explaining this is arguably the media synchronicity theory [7]. Voice offers superior feedback and symbol variety - conversants can convey more concepts, more quickly - and this is useful for coordination of urgent action. Text chat has better parallelism - more conversations can occur at the same time - and this is useful when engaging simultaneously with different people or groups in a complex game. The reprocessability of text relative to voice allows people to read back through a conversation, which is an advantage when players are distracted by gameplay activity or need to go away-from-keyboard.

Most online communicators today have a repertoire of media to choose from, as described in [25]. In their study of VoIP use in WoW, Williams et al pointed out that it was the mix of voice and text, as opposed to voice in itself, that is superior to text-only communication [26]. Users who adopt voice retain the option to use text in situations where the latter is better suited, and can fit the medium to the task, such as is discussed in [25, 29]. However, in agreement with [10] and [17], the decision to use a particular medium may be made by the team rather than the individual.

6. CONCLUSION

Our data confirm the complexity of MMORPG communication. These games combine fictional worlds and identity exploration with real people playing in real settings. User-to-user communication lies right at this intersection of reality and fantasy. The suitability of different media to MMORPGs depends on a dynamic compromise between the need for people to understand their collaborators and their desire for privacy and identity-play. Text may persist as the best option in some scenarios. Much of the experience of MMORPG players is in accordance with existing CMC theory. However some unique features of MMORPG communication will require refinement or development of theory.

7. ACKNOWLEDGMENTS

We would like to thank our study participants, the reviewers of an earlier version of this paper, and Microsoft Research Asia for supporting our research.

8. REFERENCES

1. Bartle, R. (2003) *Designing Virtual Worlds*. New Riders, Indianapolis.
2. Bartle, R. (2003) Not Yet You Fools! at Game+Girl=Advance. http://www.gamegirladvance.com/archives/2003/07/28/not_yet_you_fools.html
3. Bos, N., Olson, J., Gergle, D., Olson, G. and Wright, Z. (2002) Effects of four computer-mediated communications channels on trust development. In Proceedings of the SIGCHI conference on human factors in computing systems.
4. Carlson, J.R., George, J.F., Burgoon, J.K., Adkins, M. and White, C.H. (2004) Deception in computer-mediated communication. *Group Decision and Negotiation*: 13 (1) 5-28.
5. Carlson, J.R. and Zmud, R.W. (1999) Channel expansion theory and the experiential nature of media richness perceptions. *Academy of Management Journal*: 42 (2) 153-170.
6. Daft, R.L. and Lengel, R.H. (1986) Organizational information requirements, media richness and structural design. *Management Science*: 32 (5) 554-571.
7. Dennis, A. and Valacich, J. (1999) Rethinking media richness: Towards a theory of media synchronicity. In Proceedings of the 32nd Annual Hawaii International Conference on System Sciences.
8. Erickson, T. and Kellogg, W.A. (2000) Social translucence: An approach to designing systems that support social processes. *ACM Transactions on Computer-Human Interaction*: 7 (1) 59-83.
9. France, E.F., Anderson, A.H. and Gardner, M. (2001) The impact of status and audio conferencing technology on business meetings. *International Journal of Human-Computer Studies*: 54 (6) 857-876.
10. Fulk, J., Schmitz, J.; Steinfield C. W. (1990) A social influence model of technology use. in Fulk, J.S., C. W. ed. *Organizations and Communication Technology*, Sage Publications, Newbury Park, 117-140.
11. Gibbs, M., Benda, P. and Wadley, G. (2006) Proximity-based chat in a first person shooter: Using a novel voice communication system for online play. In Proceedings of the 3rd Australasian Conference on Interactive Entertainment.
12. Halloran, J., Fitzpatrick, G., Rogers, Y. and Marshall, P. (2004) Does it matter if you don't know who's talking? Multiplayer gaming with voiceover IP. *Conference on Human Factors in Computing Systems* 1215-1218.
13. Haythornthwaite, C. (2002) Strong, weak, and latent ties and the impact of new media. *Information Society*: 18 (5) 385-401.
14. Hew, K., Wadley, G. and Gibbs, M. (2004) Usability and sociability of the Xbox Live voice channel. In Proceedings of the First Australian Workshop on Interactive Entertainment.
15. Löber, A., Grimm, S. and Schwabe, G. (2006) Audio vs chat: Can media speed explain the differences in productivity? In Proceedings of the European Conference of Information Systems.
16. Lober, A., Schwabe, G. and Grimm, S. (2007) Audio vs. chat: The effects of group size on media choice. In Proceedings of the 40th Annual Hawaii International Conference on System Sciences.
17. Markus, M.L. (1990) Toward a "critical mass" theory of interactive media. in Fulk, J. and Steinfield, C.W. eds. *Organizations and Communication Technology*, Sage Publications, Newbury Park, 194-218.
18. Ngwenyama, O.K. and Lee, A.S. (1997) Communication richness in electronic mail: critical social theory and the contextuality of meaning. *MIS Quarterly*: 21 (2) 145-167.
19. Short, J., Williams, E. and Christie, B. (1976) *The social psychology of telecommunications*. Wiley, London.
20. Stritzke, W.G.K., Nguyen, A. and Durkin, K. (2004) Shyness and computer mediated communication: A self-presentational theory perspective. *Media Psychology*: 6 (1).
21. Turkle, S. (1995) *Life on the screen: Identity in the age of the Internet*. Simon & Schuster, New York.
22. Wadley, G., Gibbs, M. and Benda, P. (2005) Towards a framework for designing speech-based player interaction in multiplayer online games. In Proceedings of the 2nd Australasian Conference on Interactive Entertainment 2005.
23. Wadley, G., Gibbs, M., Hew, K. and Graham, C. (2003) Computer supported cooperative play, "third places" and online videogames. In Proceedings of the Australasian Computer-Human Interaction Conference.
24. Walther, J.B. (1992) Interpersonal effects in computer-mediated interaction: a relational perspective. *Human Communication Research*: 19 (1) 50-88.
25. Watson-Manheim, M.B. and Belanger, F. (2007) Communication mode repertoires: dealing with the multiplicity of media choices. *MIS Quarterly*: 31 (2) 267-293.
26. Williams, D., Caplan, S. and Xiong, L. (2007) Can you hear me now? The impact of voice in an online gaming community. *Human Communication Research*: 33 (4) 397-535.
27. Williams, D., Ducheneaut, N., Xiong, L., Zhang, Y., Yee, N. and Nickell, E. (2006) From tree house to barracks: The social life of guilds in World of Warcraft. *Games and Culture*: 1 (4) 338-361.
28. Yee, N., Bailenson, J., Urbanek, M., Chang, F. and Merget, D. (2007) The unbearable likeness of being digital: the persistence of nonverbal social norms in online virtual environments. *Cyberpsychology and Behaviour*: 10 (1) 115-121.
29. Zigurs, I. and Buckland, B.K. (1998) A theory of task/technology fit and group support systems effectiveness. *MIS Quarterly*: 22 (3) 313-334.