Towards a Framework for Designing Speech-Based Player Interaction in Multiplayer Online Games

Greg Wadley  Martin R. Gibbs  Peter Benda
Department of Information Systems  Department of Information Systems  Customer Motivation and Behaviour
The University of Melbourne  The University of Melbourne  Telstra Research Laboratories
Parkville, Vic, 3010  Parkville, Vic, 3010  Clayton, Vic, 3168
greg.wadley@unimelb.edu.au  martin.gibbs@unimelb.edu.au  peter.benda@team.telstra.com

ABSTRACT
In this paper we consider the question of how best to design voice communication for use in online multiplayer games. We propose a framework to help shape both our understanding of games and how we approach games as objects of study and as artifacts to be designed. The framework suggests a focus on designing player-to-player interaction within games. We discuss existing implementations of voice communication in multiplayer games, and the metaphors implicit in their design. We propose alternative metaphors that might be used in future designs. We conclude with an outline of work currently in progress that is attempting to understand how best to design for voice communication in Massively Multiplayer Online Role Playing Games (MMORPGs).

Categories and Subject Descriptors
H.5.2 [INFORMATION INTERFACES AND PRESENTATION]: User interfaces – User-centered design, Voice I/O.

General Terms
Design, Human Factors

Keywords
Video games, voice communication, sociability

1. INTRODUCTION
Online Multiplayer Gaming has become an important segment of the large videogames market. Many online games now come with features to enable players to talk to one another using Voice-over-IP (VoIP) and related technologies. Using voice as a means of communication is considered by many to be superior to typed text in some games. However, it has been suggested that voice will detract from immersion in online role playing games, and that text should remain the preferred communication medium in this genre of games (Bartle, 2003). Further, while some research has shown that voice can improve communication and team coordination in a fast-paced action game (Halloran et al., 2004), not all users readily adopt voice communication in online games (Wadley et al., 2003). Research suggests that this may be due to suboptimal configuration of the voice channel in some games (Gibbs et al., 2004).

In this paper we consider the question of how best to design voice communication for use in online multiplayer games. We begin by considering the frameworks typically used to analyze games. These frameworks help shape both our understanding of games and how we approach games as objects of study and as artifacts to be designed. As a result, the frameworks we use to classify and understand games are important influences on game design. Largely absent from existing classification frameworks is a means of grappling with player-to-player interaction within games. Thus, we suggest, in addition to understanding games in terms of narrative, simulation and game play, we should also understand games as technologies that provide a social experience.

Following this discussion, we present a fruitful approach to understanding games as cooperative social endeavors, which draws insights from sociologically-informed work on human-computer interaction and computer-supported cooperative work. We then move on to discuss metaphors that could be leveraged by game designers to inform their design of voice communication facilities for use in games. We conclude with an outline of work currently in progress that is attempting to understand how best to design for voice communication in Massively Multiplayer Online Role Playing Games (MMORPGs).

2. A GAME CLASSIFICATION FRAMEWORK
Several schemes for defining, analyzing and categorizing videogames have been proposed. Juul (2003) extended earlier definitions to include digital games while excluding non-games such as hypertext fiction and acknowledging borderline cases such as simulations. Aarseth et al (2003) sought to analyze the essential properties of videogames and proposed dimensions along which games can be categorized. Lindley (2003), using a similar strategy, regards videogames as systems which structure player-computer interactivity over several timescales by integrating three sub-systems: gameplay, simulation, and narrative. Particular games or genres can be classified according to the relative degree to which they embody these sub-systems. Lindley represents the subsystems as three vertices of a triangle: a game’s classification is then represented by placing a point inside the triangle, indicating the relative importance within that game of gameplay, simulation and narrative. Tetris, for example, would be placed near the ‘gameplay’ vertex, while GTA3 would be closer to ‘simulation’, and HalfLife to ‘narrative’.

With the recent popularity of games that allow geographically-dispersed people to play together via the Internet, gaming increasingly consists of social experiences mediated by communication technology. A framework for understanding videogames should accommodate these computer-mediated communications. Player interactions help to structure the gaming experience, yet are neither gameplay (though they may be part of gameplay), narrative (though players may cooperate in constructing or enacting a narrative), nor simulation (though
Communication may occur within a simulation, or employ a simulation of real-world communication technology. Therefore we propose adding a fourth subsystem, ‘player interaction’, to Lindley’s framework.

![Lindley’s classification framework, augmented](image)

This four-dimensional framework (displayed above as a tetrahedron) recognizes that social interaction can significantly impact a player’s experience, that videogames vary in the degree to which they support social interaction, and that social interaction is differently important in different videogames.

We can represent the distinction between single- and multi-player versions of the same game by moving games toward or away from the ‘player interaction’ vertex.

Largely absent from existing frameworks used to classify and understand games is a means of grappling with player-to-player interaction. This augmented framework highlights the need to consider the nature and quality of player interaction in online games, and helps us to understand and design games as technologies that provide a social experience.

### 3. COMPUTER-SUPPORTED COOPERATIVE PLAY

We have previously described videogaming over the Internet as “computer supported cooperative play” or CSCP (Wadley et al., 2003). We understand CSCP as mutual engagement by two or more individuals in recreational activity mediated by a computing environment. We chose the term because it resonates with CSCW (computer supported cooperative work), a term used to describe technologies that allow people to work collaboratively online. We believe that online games can be successful vehicles for player interaction if game designers apply the same principles that guide convivial and effective user interaction in CSCW technologies.

Previously we have used the CSCW design concepts of ‘sociability’ (Preece and Maloney-Krichmar, 2003) and ‘social translucence’ (Erickson and Kellogg, 2000) to better understand technologies for player interaction in online games. Sociability has been described as ‘planning and developing social policies and supporting social interactions’ (Preece and Maloney-Krichmar, 2003, p605), while Social Translucence emphasizes the importance of making socially significant information visible to participants in digital environments. We argued that sociability within online game environments is enhanced by making socially salient information available to players, according to interactional affordances and constraints that are sensibly designed and well understood by the players.

With the incorporation of CSCW concepts into the framework for analyzing games we are in a position to consider the design of voice communication in online games.

### 4. CURRENT IMPLEMENTATION OF VOICE COMMUNICATION IN GAMES

A recent innovation in multiplayer online games is the addition of technology that allows players to communicate with each other using VoIP rather than typed text. VoIP is available on several gaming platforms.

For example, in Microsoft’s Xbox Live network, voice is the only communication channel available. Xbox consoles do not have keyboards, and the Xbox Live subscription kit includes a headset with a microphone. Microsoft mandates that voice be implemented for player communication in all Xbox Live games. Microsoft also offers synchronous voice chat and asynchronous voice messaging to Xbox Live subscribers, outside and independent of games.

Most multiplayer games on Xbox Live are team-based first-person shooters (FPS). Voice communication seems well suited to supporting the fast-paced play and real-time tactical coordination of small teams that is necessary in these games (Halloran et al., 2004). Some Xbox Live games, such as deathmatch FPS and racing games, do not involve team coordination. In these the voice channel can be used for taunting opponents and for general chat. There are currently no massively multiplayer games for Xbox Live, though some are in development.

By contrast, microphones and headsets are not standard peripherals on personal computers, and PC game developers cannot assume that their customers will be able to use a voice channel. Few PC games provide a voice channel within the game software (CounterStrike is a recent exception). Groups of players who wish to communicate using voice in a PC game must usually agree to purchase headsets and install third-party software such as RogerWilco or TeamSpeak. These services allow groups of players to set up and administer voice channels for their use, independent of the game software they are using. This has become common practice in some FPS games. Anecdotal evidence suggests that some guilds in PC-based MMORPGs are adopting this technology.

Although typed text has been the traditional medium of communication online, speaking is a more natural way for people to communicate. VoIP seems at first glance to be ideally suited to online videogaming. It is faster than typing, and doesn’t interfere with use of a keyboard or game controller, allowing communication to occur simultaneously with avatar movement, weapon use and other activity. However when we examined use of the voice channel in some Xbox Live games, we found it to be lacking in some respects. In particular, players experienced problems of misuse of the channel, of identifying who was speaking, and of choosing to whom they wanted to speak (Gibbs et al., 2004). These problems of sociability and a lack of social translucence in the voice channel negatively impacted on player interaction in these games. This result suggests a need for researchers and designers to attend more closely to the design of the voice channel and the communication metaphors implicit in existing voice channel configurations.

---

1 Lindley (2003) extends his original three-point model in two different ways: however these extensions are made to incorporate media other than videogames. We only consider the three-point videogame framework here.
5. METAPHORS FOR DESIGNING VOICE CHANNELS IN GAMES

Existing voice channels in games are often configured in a broadcast mode, so that every utterance is heard by every player (or every team mate). This is analogous to communicating by two-way radio. Voice communication using a radio metaphor is suited to games that simulate an activity in which radio is commonly used, such as infantry combat, flight, or emergency rescue work. The core activity in many games is the negotiation of (virtual) 3D space: location, movement, finding and hiding from enemies. A successful team is one whose members coordinate effectively to deal with these features of physical space despite not being able to see each other. In games as in real life, two-way radio is an efficient way to coordinate a small but geographically-dispersed group in real time.

However radio may be an inappropriate metaphor for voice communication in games that simulate activities other than the negotiation of space (for example operating a business, building a simulation city, social maneuvering, and so on). Furthermore, just as weapons in games are not universally powerful, and would detract from game play if they were, voice technologies need not seamlessly connect all players equally or instantly. We believe that game designers need a range of metaphors, well-understood by players, for structuring the use of voice appropriately in different games. Example metaphors could include:

**Two-way radio:** Transmission is equally available to all players at all times, and all utterances are received by all players. This metaphor is well-suited to the coordination of groups of soldiers, vehicles and the like. Players may need to develop a voice protocol and radio discipline similar to that used by real radio operators. Radio communication may break down under certain circumstances (for example as team-mates disappear behind a mountain) and may be jammed by an enemy.

**Mobile telephone:** Transmission and reception are available anywhere in the virtual world; however conversations must be deliberately initiated, and utterances are transmitted between pairs of players rather than broadcast to all players. Communication may have a cost attached. The phone number of the called player must be known.

**Land-line telephone:** Communication is only available at particular places in the virtual world. Players must be near a fixed point in order to communicate. That point could be represented by a telephone, magic stone or whatever is appropriate to the game theme. Telephones normally transmit between pairs of people, but could work in a restricted-broadcast or conference-call mode.

**Physical transmission of sound in air:** When no communication technology is available, speech must be carried as sound waves in air. Communication is only available between players whose avatars are close by. The loudness and clarity of utterances can be influenced by the speaker, and attenuate with the distance between speaker and listener. These restrictions allow the speaker some control over who receives particular utterances.

In some games it could be desirable for players to be able to switch between speech metaphors according to circumstances; for example players could choose to speak either locally or by radio. The ability to use the voice channel could be a privilege to be earned through achieving game goals. Just as players have to find weapons before they can fight effectively, finding a radio or telephone before being able to speak could be part of gameplay, as could jamming enemy communication or establishing repeater stations to transmit to new areas in the game.

These metaphors for designing a voice channel would in effect be ‘social policies’ (Preece and Maloney-Krichmar, 2003), enforced by the game designer, which enhance the sociability of a voice channel by constraining the channel’s use in ways that are easily understood by players. Such policies would help ensure that the addition of VoIP technology improves rather than detracts from the experience of player interaction in videogames.

6. RESEARCH IN PROGRESS

We are investigating how voice communication can be used by players of MMORPGs. Our project involves the use of ethnographic observation, interviews and focus group sessions to develop a rich understanding of the ways in which participants can make use of voice in this genre of games.

MMORPGs are persistent virtual 3D environments in which success is difficult to achieve without cooperating within a group, and player interaction and conversation are important activities. Missions in an MMORPG are similar to FPS play in that they sometimes involve fast-paced combat and real-time coordination of tactics within a small group. However MMORPGs differ from FPS in that team formation often occurs in-game, and socializing occurs in large virtual areas, inhabited by tens or hundreds of players who do not know each other outside the game. Clearly not all players in an MMORPG could share the same voice channel. It is not clear how voice can best be configured to support this kind of interaction. Further, it has been suggested that using voice will detract from immersion and role-play in these games (Bartle, 2003). We will explore this issue; however it seems that little player-to-player communication in MMORPGs is in-character ( Tosca, 2002), and so we expect that this will not be a problem for most players.

Our project consists of two phases. Phase one is laboratory based. Small groups of two to four experienced MMORPG players are observed while playing games in a usability lab. The individuals in each group are physically isolated from one another, but able to communicate using VoIP. Following these sessions, a focus group discussion is held with the participants to discover their initial reactions to using voice in an MMORPG.

Phase two seeks to understand the adoption and appropriation of voice communication in MMORPGs over an extended period of time, using ethnographically informed data collection methods. Small groups of participants will play MMORPGs cooperatively over an extended period of three months. Participants can play the game as much as they wish, but will commit to a regular time once a week when they will play together and be observed by the research project team. Following these observation sessions, selected participants will be interviewed.

Participants will be physically separated so that only online communication (by voice or text) is possible. They will be formed into teams that cooperate to complete tasks within the game-world and compete with other teams.

At this time there are no MMORPGs available with a voice channel implemented. Therefore we are using a third-party VoIP product with PC-based MMORPGs to make voice communication available to participants in our study. With appropriate channel configuration this product will enable a
Although our research is in progress, we can speculate about which voice communication metaphors are best suited to which game genres.

Two-way radio seems to be the best way for members of a small team to coordinate tactics and achieve a goal while moving around a geographical area. Radio is especially useful if team members are sometimes distant or hidden from each other. The ability to communicate is independent of game location. (example game: CounterStrike)

The other voice communication metaphors represent restrictions of the radio metaphor, and players may only accept them if the overall benefit outweighs losing the ease and freedom of radio.

Proximity chat (using the metaphor of sound traveling in air) may be necessary where large numbers of players would clutter a shared radio channel. It may be preferable in a game where team members are normally close to each other, and the benefit of cues as to which direction a sound is traveling from outweighs the inability to speak with distant players. Currently technology is being developed to implement this communication metaphor in crowded virtual spaces (Boustead and Safaei, 2004). In games involving very large teams, players often interact mainly with team mates who are close to them, and/or form small ad-hoc fire-teams to carry out short term tasks. (example game: Battlefield 1942)

Although MMORPG quests involve tactical coordination of teams, members often stay close together to gang up on solitary NPCs. Proximity chat may be appropriate in this kind of team play. It may also be useful in towns and other gathering places where many players occupy the same space and would clog a radio channel. It will be interesting to see whether proximity chat is better suited than radio to the socializing and team formation that occurs in these public spaces. Using proximity chat, players in these spaces could use familiar real-world strategies of moving around to determine who is part of a conversation.

Current MMORPGs typically allow players to choose between a variety of text chat modes – private, vicinity, team and so on – as the situation demands. These modes can be mapped to the different voice transmission metaphors we have identified: telephone, proximity and radio, respectively. Our study will test the suitability of voice to these modes of communication.

Finally, it is worth noting that the use of voice rather than text as communication medium introduces a further kind of social interaction to online gaming. In previous research, we have observed Xbox Live players moving between two social spaces: a virtual space based on Internet communication and the physical space in which the player was located (Wadley et al 2005). Speaking rather than typing brings the game into the room, making it a spectacle for people who are physically co-located with the player. This can be a disadvantage if other people in this physical space do not want to be engaged in the game. In addition, game playing is sometimes subversive, disguised as homework or office work. Voice may not be a suitable in-game communication medium in these situations. Our research will also explore these issues.

8. ACKNOWLEDGMENTS
We would like to acknowledge Microsoft Research Asia, and the IDEA lab at the University of Melbourne for supporting this research.

9. REFERENCES