Does it Matter if You Don't Know Who's Talking? Multiplayer Gaming with Voiceover IP

John Halloran[1], Geraldine Fitzpatrick[1], Yvonne Rogers[2], Paul Marshall[1]

1. Interact Lab
Department of Informatics
University of Sussex
Brighton BN1 9QH UK
{johnhall}{geraldin}{paulma}@cogs.susx.ac.uk

2. School of Information Science
University of Indiana
1320 East 10th Street, Bloomington
Indiana IN47401 USA
yrogers@indiana.edu

ABSTRACT

Voiceover IP (VoIP) now makes it possible for people in distributed online multiplayer games to talk to each other. This might not only influence game performance, but also social interaction. However, using VoIP in multiplayer games can often make it hard to know who is talking, an issue that other researchers have found to be problematic. In a 10-week study of a fixed group of adult gamers, we found that not knowing who is talking affects game performance differently according to the type of game. In team-based war games, it can have a negative effect both on learning and coordination, but in race games, where individuals rather than teams compete, it appears generally not to matter. In contrast, the impact of not knowing who is talking on social interaction is the same regardless of game type. While the social experience can be highly enjoyable, it is difficult for gamers to get to know each other. We consider the design implications for enhancing both game performance and social interaction.

Author Keywords

Multiplayer games, voiceover IP, social interaction, qualitative study.

ACM Classification Keywords

H.5.2. [Information interfaces and presentation]: User Interfaces.

INTRODUCTION

A recent development in distributed online multiplayer computer gaming is the addition of voiceover IP (VoIP): internet-based audio conferencing. This feature allows gamers to talk to and listen to each other over a headset. In a group or team of gamers, audio is fully connected, so that everyone can talk to and hear everyone else.

We are interested in how VoIP might enhance both game performance and the ability to interact socially. Previously, when gamers wished to communicate, they needed to send text communications to each other. In fast-moving contemporary games, where gamers 'inhabit' an avatar in a 3D virtual world, both hands are necessarily occupied with the control device (e.g. keyboard or console), performing other activities (e.g. overtaking, shooting). This makes typing and sending messages hard to do. A workaround for this is to attach fixed messages to hotkeys. Thus, text implementation of communications in multiplayer games tends to reduce both the number and range of messages. This can affect coordination in team-based games, as well as how far players can get to know the real people behind the avatars [1], especially since text communications allow gamers to manipulate the identity they are presenting [2]. Being able to talk removes the logistical problems associated with sending text communications, and this may have effects on game performance. In addition, being able to hear someone's voice, a voice which may be producing a wider range of utterances than with text, may make it easier to engage in different types of communication which could reveal more of the real person. Some games producers have even claimed that being able to talk while gaming could lead to new social opportunities opening up, where meeting people online could lead to meeting them in person [3].

However, all of these possibilities may be threatened if gamers are reluctant to talk, a finding from recent research into Xbox Live[TM] multiplayer gaming with VoIP [4]. The researchers report that one reason for this is that gamers do not know who is talking, and this inhibits them. In the Xbox Live[TM] distributed multiplayer gaming study we report on here, we also found that it is often difficult for gamers to know who is talking. This was a study conducted over a 10-week period where 10 adults of mixed sex, age and experience gamed together once a week, playing either a war game, or a race game.

On the face of it, it seems odd that gamers don't know who is talking as voices are different: if the association between a voice and a gamertag (gamer's 'username') is known, knowing who is talking should be clear. However, our

observations and gamer feedback suggest that the implementation of VoIP makes it difficult to distinguish voices for a number of reasons. The audio can become distorted leading to voices sounding distant or breaking up. Also, there is little or no accompanying ambient information (e.g., what else is going on around the gamer, whether they are tapping a keyboard or shifting). Even without distortion, the loss of ambient cues has the effect of making voices of the same sex sound similar to gamers, especially where there are no distinctive accents. Even if a gamer knows someone from outside the gaming context, they can still find the voice hard to recognize in-game. This issue is exacerbated by a related issue: the breaking of the association between individual communications and gamertags. In games with text communications, messages arrive prefixed by the gamertag. Associating the same gamertag with that gamer's avatar, which is normally labeled in some way, will reveal whose communication it is. Thus a two-step association is required to find out who is speaking: message-to-gamertag, and gamertag-to-avatar. With VoIP, this is much harder to achieve, because, while the same labeling features may exist, talk is not tagged with the speaker's name. This is not an issue with VoIP *per se,* but with its implementation within games when there is no 'for free' association of gamertag and utterance.

The problem of not knowing who is talking can have design implications. We might want, for example, to ask how better voice differentiation with VoIP might be achieved; how utterances might be tagged; or how other means could be used to enable the association between gamertag and utterance to be made. However, our study suggests that not knowing who is talking often does not matter, and where it does matter, there may be alternative design implications.

METHOD
We recruited 10 adults aged 20 to 48. 7 of these were male, 3 female. These included a couple, and 2 housemates. Otherwise, the members of the group were unknown to each other. Each participant was equipped with broadband internet access, an Xbox^{TM} console, and an Xbox Live^{TM} kit – everything required to do multiplayer gaming with VoIP. The group gamed together once a week for 10 weeks at a fixed time, for 60 minutes. We provided two types of game, a war game and a race game. These game types differ in terms of both purpose and the design of voice communications. War games are collaborative and team-based. Members of a team can hear and talk to members of their own team, but not other teams. Race games are all-against-all competitions, where everyone can talk to and hear everyone else. At each session, we observed and video-recorded 2 of the 10 participants, rotating around the group so that each participant was recorded twice at a 5-week interval. We supplemented this data with interviews after observations and e-mail questionnaires following each session. The findings discussed here are based mainly on field notes, video analysis and analysis of transcripts.

FINDINGS
Contrary to the findings reported in [4], we found that not knowing who is talking did not inhibit talk, which was plentiful. Talk had different functions according to the type of game being played. In race games, talk tended to be phatic, that is, concerned with establishing a mood of sociability rather than communicating information. While talk was important for this, it was not critical in any way for game performance. This meant that not knowing who was talking tended not to be an issue. In contrast, in war games some talk was phatic, but the majority was dedicated to collaboration. Here, talk was often critical to game performance. Gamers used talk to create strategy, inform each other about aims and objectives, and request or offer assistance. Talk was also important as part of the process of learning in less experienced players. All of these things often depended on knowing who was talking, and where this could not be established problems could occur. Across both types of game, we found that gamers enjoyed the social dimension of gaming. However, this was despite not knowing much about other gamers. Even at the end of 10 weeks, gamers, in responses to e-mail questionnaire items as well as interview questions, showed that while gaming they often had difficulty remembering gamertags, recognizing voices, and perceiving the real people behind the avatars.

Using VoIP in race games
In race games individuals compete against each other by moving their 'avatar' vehicles or vehicles-and-riders around a circuit. These games are easy to learn, and fun ('sandbox games' was one gamer's characterization). The following typical short excerpt from a race game called Midtown Madness^{TM} illustrates how talk functions in these games. Here, a type of tag game is being played where one car has a stash of gold, and it is the job of others to steal it by crashing into that car. The person who manages to deliver the gold to a particular point wins. In the following excerpt (and in excerpts throughout the paper), the numbers are timings in seconds from the beginning of the excerpt and the names are gamertags which have been anonymized.

0	Buzz	Oh! Can't believe it!
5	Adder	Oh come *on*, my car!
6	Chimp	(gloating having just got the gold) aaah-haaa!
8	Adder	Oh, shit
9	Rufus	Oh he's gonna go there, he's gonna go there, isn't he
13	Hero	(to himself, as in 'be careful') E-easy!
17	Rufus	Yes! Straight on 'im!

In this excerpt, at 5 seconds, Chimp steals the gold from Adder. Every non-gold-holding car has a large arrow floating ahead of them showing them where to go to catch the gold-holder, i.e., at this point, Chimp. Rufus has Chimp in his sights. On screen, there are two other kinds of visual awareness tool: a birds-eye-view map bottom-left where a non-gold-holding car can see the gold-holding car (plus all the others); and a list of gamertags top left. When a player speaks, an animation appears next to his/her gamertag in the

list. When a car (player) approaches another, the gamertag associated with that car appears floating above it. Therefore, since voices can be associated with gamertags, and gamertags with cars, voices can be associated with cars. However, even though this information is available, it is not important, in terms of game performance, to know who is speaking. This is reflected by the finding that gamers rarely referred to others directly by name. Rufus's utterances at 13 and 17 seconds are spoken in pursuit of Chimp, but are third person, as if Rufus is speaking to everyone else, himself, or to the researcher in the room. Chimp may realize he is being followed, but by whom is irrelevant, since game performance only requires him to avoid contact. Thus, utterances tend to be spontaneous expostulations or exclamations, often concerning one's own actions and efforts, especially where something is effortful and/or disappointing (the utterances at 0, 5, 8 and 13 seconds). They can include 'crowing' - celebrating in a gloating way - e.g. Chimps utterance at 6 seconds, or Rufus's at 17. Despite being able to find out who is talking, this is not needed to take action: talk is there, rather, to create mood, energy and involvement.

Using VoIP in war games
In contrast to race games, war games require team coordination on a range of tasks, e.g. defending a submarine or stealing documents. In the following sections we will see several examples of how knowing who is talking can be very important We examine breakdowns that occurred when who was talking needed to be known but was not, and how teams developed forms of talk to make sure that, where necessary, gamers knew who was talking. We will also look at examples of coordinated action mediated by talk where knowing who is talking is not important.

Learning through knowing who is talking
In the early stages of the study, we found that the most useful learning aid for the less experienced gamers was the proximity of other players' avatars. These could be followed and their actions and speech observed in order to find out how the game works. This depends on knowing who is talking. When a person doesn't know or is unable to work out who is talking, the gaming experience can become confused and disparate.

In the first-ever gaming session with a war game (Return to Castle Wolfenstein[TM]), the two observed players were Weepy (female, 20) and Lancelot (male, 48), neither of whom had previous experience of console gaming (both had PC game experience) or VoIP. Only Lancelot had experience of war games.

At one point, Weepy found Chimp, an experienced console and war gamer (with no VoIP experience). She was able to identify him when she passed her weapon sight over his avatar. This requires deliberate action: the appearance of a gamertag is not guaranteed by proximity. She followed him to a door, where both stood stationary for some seconds.

Two utterances followed: "like shooting fish in a barrel"; and a few seconds later, "one, one more flag". Chimp had not spoken up until this point, and was the only avatar visible to Weepy. In this game, there is no list of gamertags to show who is talking. Therefore, the only resource for resolving the relationship between utterance, gamertag and avatar was game knowledge – which Weepy did not have. In fact, the first utterance refers to a killing, the second to collecting flags (the object of the game). As there was no evidence of either activity in Weepy's current physical environment, this meant it was not Chimp speaking. However, without the required game knowledge she was unable to make this inference: she did not know who, in terms of the relationship of avatar to voice, she was following, or what was being talked about. This depends on knowing who is talking.

One way these kinds of problems can be overcome is through establishing that there are only two players, yourself and another avatar, that are (a) mutually visible, and (b) speaking. At one point Lancelot heard an utterance, "Lancelot can you give us some ammo", followed by the appearance of Buzz. Lancelot made no response. Another utterance followed, "Lancelot, if you press your change weapon, that gets the ammo". 'Pressing change weapon' is a console action that alters the weapon that you are using in the game, which Lancelot then did. Lancelot then said "There's a pod, there's a pod". Buzz's response was, "OK press fire", another instruction, which Lancelot followed. The result of this was that a package of ammunition appeared at the avatar's feet. This avatar then picked it up, followed by the utterance "OK, another". This made it clear that the voice Lancelot was hearing was associated with the avatar in front of him. Thus, Lancelot was able to make the association between voice, avatar and gamertag. This association depended on there being no other voices and no other avatars to confuse the issue. Lancelot used this episode to 'glue' himself to Buzz, whom he followed for the rest of the game, asking him questions and being coached.

Acting without needing to know who is talking
For new war gamers, then, knowing who is talking is often crucial. We found, however, that with learning there is a transition from dependence on this form of disambiguation, to richer strategies for working out meaning. This transition depends on learners being integrated into a team which, as a whole, then works out these strategies. The following excerpt from the 6th week into the study shows how Weepy gets 'ammo' by interacting with all the people on her team:

0	Weepy	I need ammo anybody got some?
4	Weepy	Can anyone give me ammo?
5	Cat	It's with the health packs round by the flag I think
6	Buzz	Umm, ammo at the flag
7	Weepy	Cheers (approaches flag)

Two players, Weepy and Cat, are standing near a flag, while Buzz is nearby. Here, it is co-location, more than

knowing who is speaking, that creates meaning. Meaning is disambiguated both through game knowledge: knowing what 'ammo' and 'health packs' are; and being able to refer to a shared reference point (the flag) which is known to be in the vicinity: "round by the flag"; "at the flag". Also, Weepy has developed a strategy to overcome problems involved in not knowing who is speaking - broadcasting and waiting for responses: 'anybody'; 'anyone' – where, to achieve game objectives, it is enough to know that the person who responds is on your team.

Another way of taking action without needing to know who is speaking is to use inference to work out strategy. In one game, Lancelot observed the effectiveness of something he had not seen before – a team mate's air strike on the enemy. Lancelot knew that this is what this event was called, because it was accompanied by a shout: "air strike!". Without seeing the avatar or knowing who made the statement, Lancelot asked "How did you accomplish the air strike?". The answer was, "You select 'grenade' from your weapons and you just fire it". The point here is that it was unnecessary for either player to know who was speaking. A player called Mars launched the air strike, and was the only one to have done so. Therefore, Mars understood that a query just addressed 'you' must relate to him. The interaction accomplished shared understanding of what an air strike was and the ability to mount one.

DISCUSSION AND CONCLUSION
In this paper we have looked at what happens when online gamers are able to use a new technology, voiceover IP, to talk to each other. A mix of implementation issues and game characteristics can make knowing who is talking difficult, and this has been identified as an off-putting issue for gamers in other research. Gamers often have difficulty in identifying the relations between gamertags, avatars and voices. However, in terms of game performance, we found that this only negatively affected team games, where the less expert gamers needed to be able to make the association for learning purposes, and gamers needed to find out, for example, who was issuing a request to be assisted or met. This reflects that in team-based games, coordination is essential. Knowing who is speaking is important for coordination, but is secondary to the larger issue of working out what utterances mean with respect to the game and team strategy. This can be done using contextual information and forms of addressing and questioning where knowing who is speaking is not important. In contrast, in individual-based race games, the talk was different. Here, it is not important to know who is speaking at any time. However, the fact of being *able* to speak powerfully enhances sociability as phatic utterances establish a mood of humor and energy.

In terms of design implications, therefore, not knowing who is talking is not important for game performance in race games. The awareness tools provided by Midtown

Madness^{TM} could even be removed. For war games, however, knowing who is talking is important for coordination and learning. It could be assisted by implementing features like those found in Midtown Madness^{TM}: clearer association of gamertags to avatars, perhaps by labels that are persistently attached to avatars rather than requiring explicit actions (e.g. weapons sightings), and the addition of a list of gamertags with animations to show who is speaking at any time.

However, in terms of supporting social interaction, there may be different implications. Even where it is not necessary to know who is speaking in terms of game performance, being able to identify this could help gamers to get to know each other as real people. Therefore the awareness tools provided by Midtown Madness^{TM} should stay. However, how far people *can* get to know each other may depend on the constraints placed on talk by different kinds of game. In both types of game, talk appears to enhance the gaming experience, by raising energy in Midtown Madness^{TM}, and raising the capacity for awareness and coordination in Return to Castle Wolfenstein^{TM}. The point is, talk needs to serve these functions. Our future work will consider how far the issue of restriction on what can be talked about impacts the potential for people to get to know each other. For different types of game, this will involve more closely examining how a group's generation of shared meaning develops across time and how far this puts limits on how people can represent themselves verbally.

REFERENCES
1. Halloran, J., Rogers, Y. and Fitzpatrick, G. (2003) From text to talk: multiplayer games and voiceover IP. In *Proceedings of Level Up, First International Digital Games Conference,* 130-42.

2. Wright, T., Boria, E. and Breidenbach, P. (2000) Creative player actions in FPS online video games. *Game Studies*, International Online Journal of Games Research. Online article accessed at http://www.gamestudies.org/0202/wright/ Verified 12.01.04.

3. Everquest invokes dungeons and dragons spirit. Online article and video presentation accessed at http://www.cnn.com/2001/TECH/08/17/everquest/ Verified 12.01.04.

4. Wadley, G., Gibbs, M., Hew, K. and Graham, C. (2003) Computer supported cooperative play, 'third places' and online videogames. In *Proceedings of OzCHI 2003,* 238-241.