Player Identity Dissonance and Voice Interaction in Games

Marcus Carter, Fraser Allison, John Downs and Martin Gibbs
Microsoft Research Centre for Social NUI
Computing and Information Systems
The University of Melbourne
[marcusc][fallison][jpdowns][martin.gibbs]@unimelb.edu.au

ABSTRACT
In the past half-decade, advances in voice recognition technology and the proliferation of consumer devices like the Microsoft Kinect have seen a significant rise in the use of voice interaction in games. While the use of player-to-player voice is widespread and well-researched, the use of voice as an input is relatively unexplored. In this paper we make the argument that notions of player and avatar identity are inextricable from the successful implementation of voice interaction in games, and consequently identify opportunities for future research and design.

Author Keywords
Voice interaction, Natural User Interfaces, Player Identity, Embodiment.

ACM Classification Keywords
K.8.0 [**Personal Computing**]: General - *Games*

INTRODUCTION
In the past half-decade, advances in voice recognition technology have seen a proliferation of consumer devices and software that facilitate voice-based human-computer interaction. Embedded within smartphones, smart watches and smart homes, technologies like Apple's *Siri*, Amazon's *Echo* and Microsoft's *Cortana* facilitate novel computer interfaces in new, often more natural, contexts. Voice is consequently routinely categorized under the umbrella term "natural user interfaces" (NUI) alongside gestural interfaces and eye tracking.

While the use of player-to-player voice is widespread and well-researched [15], the use of voice as an input in games is relatively unexplored. By this, we refer to the use of the player's voice as a controller where it is typically part of a multi-modal interface with other modalities such as a keyboard and mouse, console controller or gestural tracking.

CHI PLAY 2015, October 03 - 07, 2015, London, United Kingdom

ACM 978-1-4503-3466-2/15/10…$15.00
DOI: http://dx.doi.org/10.1145/2793107.2793144

Voice interaction in games is curiously polarizing. Players routinely express frustration with its accuracy, its social awkwardness, its potential for griefing, and its inefficiency. However, some instances of voice interaction are well received, attributed with increasing flow and immersion, and voice-based interfaces are increasingly emphasized in videogame marketing. We set out to understand the dimensions of these failures and successes, aiming to look deeper at player experiences that go beyond an examination of speech recognition accuracy.

In this note, we discuss the implementations of voice in four commercial games. We argue that the successful integration of voice interaction in digital games is distinct from voice interaction in other contexts, as in games it demands consideration of the identity of the voice; that is, whose voice it is that is being recognized, and how that voice is embodied. We unpack this notion and provide design recommendations for the future design of voice interaction in games.

VOICE INTERACTION IN HCI
Player-to-player voice communication in online games and virtual worlds has received some notable attention in HCI [for a recent review, see 15]. However, voice interaction in games – where a player's voice is used as an input or controller – has not. In this section, we will briefly overview the work in HCI on voice interaction.

Voice interaction, and particularly automated speech recognition (ASR), has long been of interest in HCI as a manifestly "natural" mode of interacting with computers, although skepticism has increased over the decades as the technology has been slow to live up to optimistic early expectations [1, 10]. In recent years, research on voice recognition has focused on the technical challenges of improving recognition rates and employing ASR in alternative user interfaces for users with disabilities [1, 10].

A similar pattern is evident within studies that look at computer games specifically. A small body of research has looked at ASR in the design of games, particularly for users with motor impairments [e.g. 2, 5, 6] and for speech therapy and learning games [e.g. 9, 11]; but little investigation has been conducted on the "fun" experiential side of voice interaction. For example, Sporka et al.'s [13] study of a voice-controlled *Tetris* game identified visual feedback as an important element for the usability of the voice interface, but

did not report on the potential for voice interaction to improve or degrade a player's experience with the game.

VOICE INTERACTION IN GAMES

The history of voice interaction in video games has been significantly shaped by the console market. While the earliest examples of voice interaction in games were enabled on PCs (such as *Command: Aces of the Deep* [1995], a niche submarine simulator that allowed verbal commands), voice interaction has emerged in games in parallel with console hardware developments.

Nintendo released a Voice Recognition Unit for its N64 console, although only two games went on to use it, and only one was released in English markets; *Hey You, Pikachu!* [1998] allowed players to interact with Pikachu using simple voice commands. Sega followed suit with *Seaman* [1999], a pet simulation game narrated by Leonard Nimoy in which the player used the Dreamcast microphone to converse with a humanoid fish. The presence of enabling hardware like these drove the early development of voice interaction in games.

Sixth-generation game consoles in the early 2000s introduced online play, for which voice communication between players became important. Games in the tactical combat genre required coordination between teammates, and so in some cases even came packaged with headset microphones. As a consequence, several games in this genre including *SOCOM: U.S. Navy SEALs* [2002, and sequels] and *Tom Clancy's Rainbow Six 3* [2003] were among the first to enable players to give orders to AI teammates using voice commands.

A small number of games have experimented with voice-only interaction, such as *Tom Clancy's End War* [2007], "mic mode" in *Mario Party 6* [2004], and *Lifeline* [2003], a role-play game where the player controls the game (almost) entirely through speaking commands to characters. Among the most successful are karaoke game series such as *SingStar* [2004] and *Rock Band* [2007], which typically came packaged with dedicated stage-style microphones. In 2014, the voice-only tactical combat game *There Came an Echo* successfully raised $115,569 of funding from 3,906 backers on the crowd-funding site Kickstarter.com. In most cases voice interaction has been multi-modal, with voice as a supplement to traditional controller input. However, voice interaction in games has accelerated following the introduction of devices such as the Kinect (first released for Xbox 360 in 2010), which provided built-in voice recognition technology rather than requiring game developers to license or develop their own software, subsequently significantly lowering voice interaction development costs.

APPROACH

In order to examine the nature of voice interaction in games, we examined online discussions, reviews and "Let's Play" videos around games with voice interaction. Online discussions were found on official game forums, reddit, and via Google (using strings: [game name] and ["voice" or "voice commands" or "Kinect"]). Let's Play videos were collected with similar searches. We excluded "performative" videos (e.g. by personalities such as PewDiePie) which we expected would "play up" issues with voice commands for their audience, instead focusing on longer play through videos. Professional reviews were found via review aggregator site Metacritic, which also provided several hundred amateur reviews. We used Google Translate to examine non-English reviews to see if there were voice interaction experiences specific to accents and other languages, although this provided no additional insight. In total, for the four games discussed below, we analyzed 166 professional reviews, 2,951 amateur reviews, 84 discussion threads and 69 Let's Play videos.

Each reference to voice commands was retrieved from these data sets, and coded using open and axial coding in accordance with a constructionist grounded theory methodology. This provided us a way to obtain insight into player experiences with minimal interference, in order to understand the breadth of issues and successes associated with these new interfaces in commercially available games. The following four games represent the dominant ways in which voice interaction is being used as a multi-modal interface in games, and exemplify our argument around identity dissonance presented in the subsequent discussion.

ANALYSIS

Tomb Raider: Definitive Edition

Tomb Raider: Definitive Edition utilizes voice recognition to permit simple voice commands. These allow the player to bring up menu items (e.g. by saying "show map"), switch between weapons (e.g. "pistol" or "bow"), and pause/resume the game. In our review, we found that users raised issues around the voice interaction with regard to performance and discomfort. Performance issues included reports that the speech recognition was not reliable, and more generally, complaints that it was "faster just pressing a button". We frequently saw that speed, and subsequently improved performance, were regarded as metrics by which to evaluate the voice interface, due to its effect on the player's sense of physical mastery (a stance which reflects the values of the game's "hardcore" user base [7]). In some cases, players acknowledged that the voice configuration did improve the flow of play (e.g. changing a weapon instantly while engaged in combat).

Issues of discomfort with the voice interface were raised in both online discussions and reviews. Players frequently noted that repeatedly yelling "shotgun" at their television was "uncomfortable" and "embarrassing", and that it restricted the use of the interface to when other people were not present to be bothered by the noise. Similarly, we noted multiple accounts of the "pause" command being used by non-players to grief or control players, in a way that could presumably be used by a frustrated parent to end play:

> [my] wife hates when I game with her [at] home or awake and thinks it's fun to use voice commands to turn it off and so do my kids

Splinter Cell: Blacklist

In *Splinter Cell: Blacklist* [2013] the protagonist (Sam Fisher) must navigate through areas patrolled by hostile enemy guards, using stealth rather than brute force. The player can use actions such as throwing a rock to make the enemy investigate the noise, so that Sam Fisher may ambush them from behind or sneak past undetected. In the Xbox One version of the game, the user can yell "Hey you!" to the Kinect sensor, and Sam Fisher accordingly calls out "Hey you!" in the game, making a virtual sound which the enemy guards will investigate (see Figure 2).

This feature was very well received, with various reviewers noting "the ability to relate directly with fictional characters is an [sic] powerful idea" [14] and online discussions lamenting that there were so few commands that worked with the interface. In comparison to other examples of voice interaction, players liked that they were doing what their character would actually do, rather than something "unnatural" that they would not normally say out loud (such as yelling "shotgun" in *Tomb Raider*).

Figure 2: Splinter Cell: Blacklist (2014) allows players to distract in-game enemies by shouting "Hey you!"

Issues with discomfort were not entirely absent, however, as there remained a disjuncture between actions appropriate in the game world and actions appropriate in the real-world context. One player mentioned that they had been "caught" by their partner yelling various words at their television with no feedback, in an attempt to test the game for other voice commands. We also noted that in Let's Play videos, users attempted to engage the voice recognition function with "hey buddy!" and "come here!" until the correct "hey you!" registered.

FIFA 2014

FIFA 2014 [2014; from herein just *FIFA*] is a soccer simulation game that uses a wide variety of voice commands during offline matches. The user can select substitutions (by saying "substitution" followed by the substitute's name), change team formations (e.g. "formation two"), use custom tactics (e.g. "offside trap"), and change the mentality of the players (e.g. "ultra attacking"). It is possible to do all of these things using the controller, but voice allows them without pausing play, and so avoids interrupting the experience.

The implementation of voice in *FIFA* is one of the most commended implementations in a contemporary game. We noted a large number of positive comments about the "well-conceived", "effective" and "useful" voice interactions, and Let's Play videos positively featuring the voice commands were the most numerous and watched out of the games we examined in this study. Players praised the voice interaction options both for improving their ability to perform in the game (as doing the same tasks with the controller required "pausing the game or pressing difficult button combinations on your D-Pad and los[ing] focus on the ball"), and for avoiding a sense of discomfort; a common sentiment was that the commands "don't feel artificial or put on" and were "natural".

Ryse: Son of Rome

Ryse: Son of Rome [2013] is a third-person combat game for Xbox One in which the user plays a Roman centurion, occasionally commanding other troops in battle. *Ryse* features voice commands such as "fire volley" and "charge" that are relevant to events in the game's linear story, and the opportunity to use them is triggered by in-game events.

Overwhelmingly, players spoke positively about the voice commands, with the feature commonly being referred to as "immersive", and negative comments limited to the infrequent opportunities to use them. In the context of the game's ancient Roman setting, we identified numerous instances where players "Put on the roman soldier epic voice for it and everything", reflecting the virtually embodied "real" voice we noted in the example of *Splinter Cell*.

DISCUSSION

These four games, and our corresponding analysis, overview how multi-modal voice interaction has been integrated in early eighth-generation console games. While issues remain around the accuracy of voice recognition technology and whether the implementation improves the player's in-game performance, we argue that the key issue with regard to voice interaction in games is best understood through the lens of identity dissonance.

Carter et al. [3, 4] distinguish between four types of identities present in a game play situation: the user (the "real" human who plays); the player (a social identity); the character (an identity within a game's imaginary); and the avatar (the character's virtual depiction). This framework does not suggest that players necessarily identify with their characters, but instead establishes them as separate identity constructs which may overlap and inform each other in a game-play situation.

Through this lens, it becomes clear that in the example of *Splinter Cell*, voice was well received because of a voice-

based resonance between the user's player identity and the character identity of Sam Fischer; the user saying "hey you" in the real world meant that their character said "hey you" in the virtual world, with the expected effect. Virtually embodying the player's real voice increases (at least the perception of) the overlap between the player and character identities. Players' comments indicated that this convergence of identities could be contributing to an increase in their sense of flow and immersion. Perhaps the most extreme example of this overlap is found in karaoke games such as *SingStar* [2004], in which the character in the game space is almost completely defined by the singing voice of the player.

Contrastingly, voice interaction in *Tomb Raider* afforded no such convergence, as the in-game character did not (and would not) yell "shotgun" or "reload" in the middle of combat. Cited by numerous players and reviewers as "unnatural" and "uncomfortable", we argue that this configuration causes a dissonance between the player and character identities which could diminish the player's sense of flow and immersion in the game. While in some cases changing weapons by voice command was faster (assumedly improving the flow of the experience), the identity dissonance appears to negate this positive effect.

What is interesting about approaching voice interaction in games through this lens is the way it reveals the character identity in sports simulation games like *FIFA*. One interpretation of the player's role in *FIFA* is as the manager or coach, particularly in career mode. The voice commands as implemented in *FIFA* accord with this personification; mentalities like "defensive" and tactics like "offside trap" are commands that a coach or manager might yell out from the sidelines to their players, and several commenters felt these were things they would already yell at their TV during intense moments of play. Reflecting and playing with this idea, the player's character can receive a letter from the board of directors in *FIFA* chastising them for swearing too much where the microphone could hear them.

As noted earlier, we also identified how many players would mimic the voice and (British) accent of the protagonist in *Ryse: Son of Rome* when giving voice commands to other troops. Further, rather than simply enunciating "fire volley" in a calm and reliably recognizable tone, many players would shout the command as if the urgency in their own voice would be conveyed to the virtual archers. An issue with this implementation that we note in this context (but we did not identify online) is that when the user says, for example, "Soldiers, move out!", the in-game character then actually says (in one case) "Move! Move! Keep out of the blast area!" potentially reducing the immersive effect of a virtually embodied real voice. However, these emergent practices further reflect players' desire for resonance between their own vocal identity and that of the character, and their incorporation into game design could further improve player experience around voice interaction.

This raises a risk of identity dissonance when it is difficult for the player to make their voice mimic their character's voice, impacting player experience. As the majority of games employ only male protagonists, this may potentially mean that female players will have a different, less immersive or more uncomfortable experience using voice interaction in games. This critique is particularly interesting in the context of (infamous [see 8]) comments by *Tomb Raider* executive producer Ron Rosenberg, who suggested that "when people play Lara, they don't really project themselves into the character... they're more like, 'I want to protect her'" [12]. This prejudicial, male-oriented and dissociated conceptualization of the player-character relationship (the player is "kind of like her helper", according to Rosenberg) is reflected in the configuration of voice interaction, where player-voice is configured as a command to the character rather than a convergence between the two identities present in the play situation.

These examples demonstrate how voice interaction in games with persistent identities must take into account the game's imaginary, and the identity of the player in that imaginary. Where voice interaction is not related to the virtually embodied experience, it causes dissonance between the user and their character, thereby negatively affecting the way the game is experienced and reviewed. User commentary indicates that embodying a player's voice through their in-world character provides an opportunity to increase immersion and flow, and appears to circumvent the widespread criticisms of voice interaction as "unnatural", "forced" or "embarrassing".

FUTURE WORK

From our initial exploration of existing practices, in this short paper we have contributed a theoretical understanding that can guide the design and future research into voice interaction in digital games. Further testing of this theoretical understanding is necessary, such as through experimental game design or a more in-depth study of players. Considering the wide range of contemporary commercial games that utilize voice interaction in some form, and the lack of HCI research into the usability and design of voice interaction in games, such work seems immediately necessary.

In addition, as voice and other natural user interfaces (such as gesture and eye tracking) are increasingly being integrated with modern games, understanding how player identity and embodiment influences the experience of novel interfaces may have application in other and future domains. Further, questions around voice and embodiment in games with more complex player identities (such as *SimCity*) need to be further explored.

REFERENCES

1. Aylett, M.P., Kristensson, P.O., Whittaker, S. & Vazquez-Alvarez, Y. (2014). None of a CHInd: Relationship counselling for HCI and speech technology. In *Proc. CHI'14 EA* (pp. 749-760). ACM Press.

2. Bilmes, J. A. et al. (2005). The Vocal Joystick: A voice-based human-computer interface for individuals with motor impairments. In *Proc. Human Language Technology and Empirical Methods in NLP* (pp. 995-1002). Association for Computational Linguistics.
3. Carter, M., Gibbs, M., & Arnold, M. (2012). Avatars, characters, players and users: Multiple identities at/in play. In *Proc. OzCHI 2011* (pp. 68-71). ACM Press.
4. Carter, M., Gibbs, M. & Wadley, G. (2013). Death and dying in DayZ. In *Proc. 9th International Conference on Interactive Entertainment* (article no. 22). ACM Press.
5. Flynn, S. M., & Lange, B. S. (2010, August). Games for rehabilitation: The voice of the players. In *Intl. Conf. Disability, Virtual Reality & Associated Technologies (ICDVRAT 2010)* (pp. 185-194).
6. Harada, S. et al. (2011). Voice games: Investigation into the use of non-speech voice input for making computer games more accessible. In *Proc. INTERACT 2011* (pp. 11-29). Springer Berlin Heidelberg.
7. Kirkpatrick, G. (2012). Constitutive Tensions of Gaming's Field: UK gaming magazines and the formation of gaming culture 1981-1995. *Game Studies*, 12(1).
8. Layne, A. & Blackmon, S. (2013). Self-Saving Princess: Feminism and post-play narrative modding. *A Journal of Gender, New Media and Technology* 2, http://adanewmedia.org/2013/6/issue2-layne-blackmon/.
9. Loaiza, D. et al. (2013). A video game prototype for speech rehabilitation. In *Proc. VS-GAMES 2013* (pp. 1-4). IEEE.
10. Munteanu, C. et al. (2013). We need to talk: HCI and the delicate topic of spoken language interaction. In *Proc. CHI'13 EA* (pp. 2459-2464). ACM Press.
11. Navarro-Newball, A.A. et al. (2014). Talking to Teo: Video game supported speech therapy. *Entertainment Computing*, 5(4), 401-412.
12. Schreier, J. (2012) You'll 'Want to Protect' The New, Less Curvy Lara Croft. *Kotaku*. http://kotaku.com/5917400/youll-want-to-protect-the-new-less-curvy-lara-croft.
13. Sporka, A. et al. (2006) Non-speech input and speech recognition for real-time control of computer games. In *ASSETS'06* (pp. 213-220). ACM Press.
14. Thomsen, M. (2012) New Splinter Cell: Blacklist lets players yell at guards with Kinect. *Kill Screen*. http://killscreendaily.com/articles/new-splinter-cell-blacklist-lets-players-yell-guards-kinect.
15. Wadley, G. et al. (2014). Voice in virtual worlds: The design, use and influence of voice chat in online play. *Human–Computer Interaction*, doi:10.1080/07370024.2014.987346