Word Play: A History of Voice Interaction in Digital Games

Fraser Allison[1], Marcus Carter[2] and Martin Gibbs[3]

Abstract

The use of voice interaction in digital games has a long and varied history of experimentation, but has never achieved sustained, widespread success. In this article, we review the history of voice interaction in digital games from a media archaeology perspective. Through detailed examination of publicly available information, we have identified and classified all games that feature some form of voice interaction and have received a public release. Our analysis shows that the use of voice interaction in digital games has followed a tidal pattern; rising and falling in seven distinct phases in response to new platforms and enabling technologies. We note characteristic differences in the way Japanese and Western game developers have used voice interaction to create different types of relationships between players and in-game characters. Finally, we discuss the implications for game design and scholarship in light of the increasing ubiquity of voice interaction systems.

Keywords

Voice interaction, speech recognition, media history, game design, media archaeology

[1] Interaction Design Lab, School of Computing and Information Systems, The University of Melbourne, Parkville, VIC, Australia
[2] Department of Media and Communications, The University of Sydney, Camperdown, NSW, Australia
[3] Interaction Design Lab, School of Computing and Information Systems, The University of Melbourne, Parkville, VIC, Australia

Introduction

A person speaks casually to a computer opponent over a game of chess. The computer answers in a synthesised but seemingly intelligent voice; they exchange witticisms and debate recent events. Such scenes have long been predicted by both computer scientists and science fiction writers. For decades, we have imagined that a world of playful, talkative, genuinely intelligent AI is close on the horizon - but like a mirage, it has seemed to melt away at our approach, as each step forward has revealed unanticipated challenges and technical setbacks (Huang, Baker, & Reddy, 2014; Munteanu et al., 2013). This pattern of "false dawns" (Aylett, Kristensson, Whittaker, & Vazquez-Alvarez, 2014, p. 754) and disillusionment has led to a period of neglect for the study of voice interaction (Munteanu et al., 2013).

Three parallel developments have given voice interfaces renewed relevance, particularly in the context of game design. The first is a series of rapid improvements in speech recognition, driven by increased computational power, cloud computing services, larger datasets and improved machine learning techniques (Huang et al., 2014). The second is a growing presence of conversational AI characters in technology systems outside of digital games, sometimes modelled directly on game characters, such as Microsoft's Cortana (Warren, 2014). The third is a rapid proliferation of microphones and speech processing systems in consumer gaming devices, from all-purpose devices such as smartphones and home computers to more specialist gaming equipment such as the Microsoft Kinect. These developments point to a growing opportunity for game designers to build novel gameplay experiences around voice interaction[4], and for voice

[4] Note that we use the term *voice interaction* exclusively for the use of voice to interact with technology, as distinct from *voice communication* or *voice chat*, which refer to online conversation between humans.

interface designers in other fields to learn how games have made spoken interaction with virtual characters enjoyable and engaging for users (Luger & Sellen, 2016, p. 1). The history of digital games provides guidelines for both of these perspectives.

In this article, we review the history of voice interaction in digital games from its origins in computer science research in the 1970s to the present day. The patterns of experimentation, popularity and commercial failure seen in this time period offer clear indications about how certain interaction styles are received by users, in a context where a positive user experience is the overriding consideration. We identify seven distinct phases, chronologically overlapping but distinguishable by their content, design focus and geographic centres of production. We find that the development of voice interaction games has been highly reactive to changes in the technological platforms that have enabled digital gaming, and that hardware has both shaped and constrained the design space of voice interaction games. We also identify regional differences in voice interaction game design, and compare the characteristic style of games designed in Japan to those designed in North America and Europe. Finally, we argue that the current phase of voice interaction game design is qualitatively different to those that have come before, with a proliferation of small and independent developers creating low-budget games that have a tighter focus on the affordances of the voice interface, suggesting that voice interaction game design may be in a signal period of experimentation and development of the form.

Data and Method

Histories of speech recognition have been written by Huang et al. (2014), Rabiner and Juang (2008), Baker et al. (2009) and Furui (2005). In each of these, the focus is firmly on describing

developments in the underlying technologies that enable speech recognition systems. However, as several scholars have noted, there has been a comparative lack of research on the interface design and user applications of voice interfaces (Aylett et al., 2014; Luger & Sellen, 2016, p. 1; Munteanu et al., 2013). Rather than rehash the contents of existing histories by explaining in detail the technology that enables voice interaction in games, we focus in this article on reporting the applications for which voice interfaces have been used in games, from a player's perspective. This is intended to complement the existing literature that provides a more technically-oriented perspective. However, where relevant we refer to aspects of the technology that are informative for understanding the allowances and limitations of a particular voice interaction system.

To reconstruct the history of voice interaction in digital games, we set out to archive and categorise every digital game that has received a public release and features voice input as a mode of interaction. To achieve this, we conducted searches on related keywords within selected publications in game studies and human-computer interaction, before broadening our search to general academic databases, Google Scholar, and the wider web. We searched for keywords on the game review aggregator site Metacritic, the digital distribution platform Steam, game-related wikis such as Giant Bomb, discussion forums such as Reddit, and enthusiast online game catalogues such as the Video Game Console Library. Additional information about each game was compiled from voice interaction research papers, platform studies texts, historical books, magazine articles, news reports, game reviews, published interviews, sales data, marketing materials, publisher websites and the games themselves. Wherever possible, findings were checked with multiple sources. We used the automated translation service Google Translate to review websites in languages other than English; however, this remains a study primarily

informed by English-language sources.

In our analysis of these materials, we have sought to classify design patterns and trace connections between games that employ similar styles of voice interaction. According to Suominen's (2016) categorisation of game histories, our approach is primarily genealogical, as we map out the lineages and subcategories that have emerged among voice interaction games, and to some extent pathological, as we seek to explain these developments by reference to their conceptual antecedents and material preconditions. In a loose sense we are engaged in media archaeology (Apperley & Parikka, 2015; Parikka & Suominen, 2006), although within the constraints of this article we can account only briefly for the broader media-cultural contexts in which voice interaction games have developed.

Antecedents to Digital Voice Interaction Games

The idea of conversing with inorganic objects is far older than the digital age. Around the 3rd Century BCE, the Greek sophist Callistratus wrote of a statue of an African king that was able to speak:

> There was in Ethiopia an image of Memnon, the son of Tithonus, made of marble; however, stone though it was, it did not abide within its proper limits nor endure the silence imposed on it by nature, but stone though it was it had the power of speech. (Fairbanks, 1931, p. 407)

Speech recognition technology itself predates the digital computer by three decades, in the form

of children's toys. The first such device was a grumpy-looking toy bulldog (Figure 1), variously known as Radio Rex or the "Wireless Pup", that sprang out of its kennel when its name was called (Cohen, Giangola, & Balogh, 2004, p. 15; 'Toys that obey your voice', 1916, p. 718; Gernsback, 1916, p. 104). First patented in 1916 and refined in subsequent years (Berger, 1916, 1917, 1918), Rex worked on a mechanical form of proto-speech recognition: the lever that pushed Rex forward was held back by an electromagnet, whose power connection was tuned to disconnect when vibrated at around 500Hz. This matches the frequency of the vowel sound in "Rex" - ironically calibrated to the vocal pitch of an average adult male rather than a child.

With the advent of computers, machines that could think and speak entered the realm of plausibility, and stories of humans matching wits and words against artificial intelligence quickly became a staple of science fiction. In the 1960s, both *2001: A Space Odyssey* and *Star Trek* showed audiences a futuristic spaceship AI that addressed its crew by voice and played chess against them, only a few years before voice-controlled computer chess would become a reality.

Research Games

In 1973, Raj Reddy, Lee Erman and Richard Neely at Carnegie Mellon University published a description of *Voice-Chess*, a digital chess game built on the Hearsay-I speech recognition system (Reddy, Erman, & Neely, 1973). Like the AIs of *Star Trek* and *2001: A Space Odyssey*, *Voice-Chess* could recognise standard chess instructions such as "Bishop to Queen Three" and update the on-screen board accordingly. This is perhaps the earliest example of a prototype speech recognition system that used a game as its use case. The turn-taking structure and well-defined vocabulary of chess moves provided a template for utterances that were constrained and

standardised enough for Hearsay-I to process.

The development of Hearsay-I was funded by the United States Advanced Research Projects Agency (ARPA) Speech Understanding Research program (Lea & Shoup, 1979; Rabiner & Juang, 2008). Several other pioneering systems were developed under this program, including Hearsay-II and Harpy at CMU, and Hear What I Mean at Bolt Beranek and Newman. Kai-Fu Lee at CMU would later combine elements of Harpy with improved statistical methods to create the first Sphinx system (Lee, 1988), whose successors continue to be among the mostly widely used tools for speech recognition today, including in digital games such as *In Verbis Virtus* (2015).

As speech recognition technology has improved, researchers have continued to use digital games as prototypes or research objects, particularly when exploring novel interaction methods (e.g. Dow, Mehta, Harmon, MacIntyre, & Mateas, 2007; Nanjo, Mikami, Kunimatsu, Kawano, & Nishiura, 2009). One notable branch of human-computer interaction research has explored voice input as an alternative control modality for users with motor impairments (Harada, Wobbrock, & Landay, 2011; Mustaquim, 2013), sometimes using non-verbal voice actions such as volume or pitch rather than words (Harada et al., 2011; Igarashi & Hughes, 2001; Sporka, Kurniawan, Mahmud, & Slavík, 2006).

False Starts

By the early 1980s, speech recognition technology was being added to mass market products. The Milton Bradley Company was first to bring voice interaction to the commercial digital

games market. In 1983 it released the "MBX" module for Texas Instruments' TI-99/4A computer, which added speech recognition and speech synthesis capabilities (Herman, 1997, p. 282; Mace, 1983, p. 27). Milton Bradley published seven games that took advantage of the speech recognition feature, including *Championship Baseball*, in which players could tell their baseball team where to throw the ball. Seeing this, Atari commissioned Milton Bradley to develop a "Voice Commander" module for the Atari 2600 console (Herman, 1997, p. 92).

An even more ambitious project was the "Halcyon" game console (Figure 2), created by Rick Dyer's RDI Video Systems, developer of the popular arcade game *Dragon's Lair*. Unlike the Milton Bradley add-ons, Halcyon was a whole console dedicated to voice input, and designed to be controllable entirely through spoken commands. Marketing for the Halcyon proclaimed it to be "the first and ONLY home entertainment system featuring speech synthesis, voice recognition, and artificial intelligence" (Kinder & Hallock, n.d.), and in interviews Dyer described the system as being similar to the artificial intelligence HAL 9000 from *2001: A Space Odyssey* ('Lasers and Computers', 1985). The console contained a dedicated speech recognition subsystem, which could recognise a player saying their own name, and had the capacity to learn more than one thousand words. Commands aside from "yes" and "no" had to be pre-trained by the player before the system could recognise them, and even then the error rate was frustratingly high ('Lasers and Computers', 1985; Dark Watcher & 98PaceCar, n.d.), but nevertheless this technology was highly advanced for a consumer product at the time.

All three platforms fell victim to bad timing. In 1982 the markets for both home computers and videogame consoles had experienced a glut, which by 1983 had spilled over into a costly price

war. A flood of cheap game hardware and poor-quality games caused a sharp drop in profits for videogame companies and a decline in consumer interest, which led to the industry-wide North American videogame crash of 1983 (Wolf, 2008, p. 105). Texas Instruments withdrew from the home computer business, dooming the MBX module a month after it began production (Mace, 1983, pp. 22–27). Atari shelved the Voice Commander indefinitely, prompting Milton Bradley to bring a lawsuit against it for breach of contract (Herman, 1997, p. 92). Halcyon was launched in 1984 with only two games and a price tag of more than US$2,000; sales were non-existent, and RDI Video Systems went bankrupt almost immediately (Wolf, 2008, p. 101; Wolf, 2012, pp. 353–354).

At the same time, however, Japan was beginning its rise to dominance in videogame production. In 1983, Nintendo released the Family Computer (Famicom) console to unprecedented success (Altice, 2015, p. 200). The Famicom included voice interaction almost as an afterthought: it came with two controllers, one of which had a built-in microphone that could amplify the player's voice through the television speaker. The audio processing system was basic, capable only of registering sound as a binary on-or-off signal (Altice, 2015, p. 25). For the first two years of the Famicom's release, no games made use of its microphone; but from 1985 a number of developers began to add voice interaction elements to their games, usually in the form of an optional or secret ability rather than a core game mechanic. Given the extreme limitations of the hardware, Famicom games show a surprisingly broad range of voice actions: from shouting to kill enemies (*The Legend of Zelda*), to blowing air to spin a roulette wheel (*Kamen Rider Club*), to "bargaining" with a merchant (*Kid Icarus*).

In 1987, the first dedicated karaoke videogame was published on the Famicom: *Karaoke Studio*. The cartridge for *Karaoke Studio* included a physical attachment for a stage-style microphone. The player sang into this microphone while music played and song lyrics appeared on screen, accompanied by an animated video. The player's singing was amplified through the television speakers, and given a score at the end of the song; although this was based on little more than timing, and as with later karaoke games there was no element of speech recognition ('Karaoke Studio', n.d.). It would be ten years until the next karaoke videogame was released, but since that time the genre has continued to follow the basic form laid down by *Karaoke Studio*.

Despite its novelty, voice interaction was never seen as a big part of the Famicom's success. Later versions of the console, including the Nintendo Entertainment System (NES) that was sold outside Japan, removed the microphone, and the flow of games on the system that used the feature gradually dried up (Altice, 2015, p. 26). It would be a number of years before Nintendo, or any other company, reintroduced voice interaction technology to a games console.

Digital Companions and the DS in Japan

The success of the Famicom/NES helped to revive the videogame industry, and moved its centre of gravity from the United States to Japan. Nintendo, Sega and Sony dominated the game console market throughout the 1990s, and they defined the capabilities of their respective game consoles in terms of computing power. The early and late stages of the decade were commonly referred to as the "16-bit era" and "32-bit era" in games, reflecting the popular focus on CPU size at the time. None of these consoles featured a built-in microphone, and consequently no voice interaction games were released for those platforms.

In 1998, Nintendo again brought voice interaction back to console gaming with the Voice Recognition Unit, an add-on to the Nintendo 64 console that paired a microphone with a speech recognition pod that could encode the player's voice inputs (Provo, 2000). However, only two games made use of the unit - 1998's *Hey You, Pikachu!* and 1999's *Densha De Go! 64* - and only the former was sold outside Japan. Although Nintendo quickly abandoned the Voice Recognition Unit, *Hey You, Pikachu!* earns a significant place in this history as the first "virtual pet" game to use voice interaction.[5] The game allows players to befriend and converse with the title character, a friendly Pokémon, while controlling a relatively nondescript child avatar. This design pattern was repeated in several games by Japanese developers over the following few years: *Seaman* on the Sega Dreamcast (Figure 3), *Lifeline* on the Sony PlayStation 2 and *N.U.D.E.@* on the Microsoft Xbox. Like *Hey You, Pikachu!*, these games use a first-person perspective, but place the narrative focus on a second-person "friend" or "pet", like Pikachu. In all four games, the player's main task is to converse with the virtual companion to teach them or guide them through the gameplay scenarios, as the relationship between the characters develops over time. This relationship-oriented voice interaction has remained a characteristic of Japanese game design and a rarity in Western game design for the next decade.

Yet another configuration of voice interaction game design appeared on the PlayStation 2 in 2001. *Yoake no Mariko* positions the player as an actor in various B-grade films, and presents

[5] *Apple Town Monogatari*, a life simulation game for the Famicom, is arguably an earlier case, although its voice interaction is limited to two hidden "Easter eggs": by shouting at the right moment, the player can make their virtual friend fall off a ladder, or cause a chirping bird to appear.

them with an on-screen script including stage directions for each line's delivery. The player is scored on how well they perform their lines, leading to an experience somewhere between karaoke and role-play. Although not an especially popular or influential game, *Yoake no Mariko* is notable as an example of the use of voice input to create a distinct new game genre.

The fact that this crop of voice interaction games emerged within the space of a few years after the long winter of the 1990s was not a coincidence, but neither was it driven by a sudden surge in consumer demand for voice interaction. Rather, it was a response to the hardware. This was the era in which consoles began to allow online multiplayer gaming, beginning with the Sega Dreamcast. Online voice chat was a fast and socially engaging way for players to communicate without taking hands off the controller (Wadley, Carter, & Gibbs, 2015, p. 360), and so each of the major game platforms received a microphone peripheral that players could purchase, often bundled with an online game. Once microphones had become common, voice interaction became a much more cost-effective prospect for game developers to explore, as consumers no longer needed to pay for a dedicated microphone for the sake of just one or two games. Conversely, a voice interaction game that came bundled with a microphone, such as *Seaman*, was a more attractive prospect to consumers as the microphone could be used for online voice chat after the game itself was done with.

Meanwhile, Nintendo continued to experiment with voice interaction systems, separate from any online voice chat applications. Several games on the Nintendo GameCube were bundled with a microphone to allow voice control, including the bizarre *Odama*, in which the player commands the movement of feudal Japanese soldiers to avoid being crushed by an enormous mystical

pinball. This culminated in the system that brought voice interaction gaming to its largest audience yet: the Nintendo DS.

The Nintendo DS was the successor to Nintendo's Game Boy series, and it extended the handheld console format in a number of ways. Most prominent were its two LCD screens, including a touchscreen and stylus. It was also the first game console since the Famicom to include a built-in microphone. Game developers responded to this unorthodox set of features with new and unusual game mechanics. *Brain Age*, one of the console's flagship titles, tests the player's mental agility with rapid-fire colour naming. *Nintendogs* lets players teach tricks to a virtual dog by calling commands like "sit" or "roll over". The *Phoenix Wright* series, about the trials of a cartoon lawyer, allows players to interrupt court proceedings with a shout of "Objection!" *The Legend of Zelda: Phantom Hourglass* asks the player to extinguish torches and spin windmills by blowing into the microphone. Each of these games was critically and commercially successful, and the DS itself sold more than 150 million units worldwide, rivalling the PlayStation 2 as the highest-selling game console of all time (Lynch, 2013). This brought voice interaction gaming to a much wider audience than had ever experienced it before, in a variety of styles. Speech recognition was not yet a seamless technology, however, and the novelty of voice inputs was offset by the console's tendency to misunderstand or fail to respond to the player's commands.

Command and Control in the West

In North America and Europe, voice interaction gaming took a long time to recover from the crash of the 1983. When it did return, it took on a distinctly different character than the style of

voice interaction games that were being made in Japan.

The first commercial videogame on the PC to feature built-in voice interaction was *Command: Aces of the Deep* in 1995. In this World War II submarine simulator, the player takes on the role of a German U-boat captain, and can speak orders to their crew through a microphone. Although *Aces of the Deep* was well received, the voice recognition was criticised for being unreliable (‘150 Best Games of All Time’, 1996, p. 74; Stevens, 1999). It proved to be the last game in the *Aces* series, and for several years it remained the only major PC game to include voice interaction. Unlike developers for the original Famicom or later platforms, PC game developers at the time could not rely on their consumers to have a microphone that could take advantage of any voice interaction features.

In retrospect, however, *Command: Aces of the Deep* clearly foreshadows the style of voice interaction that was to become characteristic of Western game design in the decade to come. Firstly, it puts the player in the role of a military authority figure in a combat situation, who barks orders at a team of underlings. Secondly, it presents a fundamentally top-down communication structure, in which the player’s subordinates are semi-autonomous agents of the player’s strategic will, rather than fully-fledged characters with inner lives and thoughts of their own to discuss. Thirdly, it affords voice control as an option that can easily be replaced with mouse clicks and hotkeys, rather than a central and indispensable part of the game’s design. This impersonal command-and-control style is indicative of the common patterns of voice interaction that came to predominate in games made by North American and European developers. Games made in Japan, by contrast, often employed a more conversational style that placed the virtual

character on a more even level with the player, and involved peaceful rather than combat-based scenarios.

The new wave of voice interaction games began to arise in North America after headset microphones were released for the Sony PlayStation 2 and Microsoft Xbox consoles. These headsets were primarily designed to allow voice chat between players in online games, but were immediately used for voice interaction. The first game to use the PlayStation 2 headset was *SOCOM: U.S. Navy Seals*, a tactical third-person perspective shooter, which included an option for the player to give verbal orders to the squad members under their command. This model was repeated in four more games in the *SOCOM* series, along with a string of other squad-based shooters: *Rainbow Six 3* and *Rainbow Six Vegas*, *SWAT: Global Strike Team*, *Greg Hastings Tournament Paintball Max'd* and, on the PC, *Unreal Tournament 2004*. In each case, the player was positioned as the leader of a combat team who could issue instructions to their semi-autonomous and largely characterless teammates.

The tactical combat command trend culminated in *Tom Clancy's EndWar*, the first game to attempt real-time strategy on a console using voice commands. Unlike the squad games that came before it, voice command was central to the design of *EndWar*. Although it could be controlled using the gamepad, the game encouraged voice input as the primary method of control. *EndWar* featured an unusually flexible speech recognition system that could understand multi-part command strings in a Who-What-Where pattern, such as "unit two, attack hostile one" or "unit four and all gunships, create group and move to target". Reviewers praised the "superbly implemented voice-recognition system" (Reed, 2008) for providing an elegant and reliable

solution to the ergonomic constraints of the console format. Unfortunately, the non-voice aspects of the game were criticised for lacking strategic depth and narrative interest. *EndWar* attracted a small cult following, but ultimately entered the annals of gaming as a curio rather than a way forward.

Following the failure of *EndWar* to find a large audience, voice interaction petered out in most game genres. However, one category bucked the trend, and by itself took voice interaction gaming to its greatest ever heights of popularity: karaoke.

The Karaoke Boom

Karaoke has been far and away the most successful genre of voice interaction gaming in its history. The first karaoke videogames appeared on the Famicom, first in the form of a mini-game within the action-adventure title *Takeshi No Chōsenjō*, and later as a fully developed game in *Karaoke Studio*, described above. The genre fell dormant in the 1990s, but with the revival of voice interaction gaming at the turn of the millenium, karaoke began a ten-year run as one of the most popular game genres of the era. Karaoke was a centrepiece of the "party game" style at the same time as videogames began to shake off their children's-toy image and achieve popularity with older demographics. Karaoke games became a feature of adult social events as a social bonding activity, often accompanied by alcohol (Fletcher & Light, 2011).

Once again, the revival began in Japan before it reached the rest of the world. The short-lived *Dream Audition* series, beginning on the PlayStation 2 in 2000, followed the basic template laid down in *Karaoke Studio*: players were given a stage-style microphone and a list of well-known

pop songs, and were scored based on their ability to hit the right notes as they sang along to an instrumental version of the song. Lyrics appeared on screen in time to the music, although as the system tracked only vocal pitch and not phonemes, the player was free to sing whatever words (or sounds) they chose. In 2003, *Karaoke Revolution* took up the format and added a piano roll-style pitch indicator for the vocals (Figure 4). It depicted the player as a character singing to a crowd, and framed the scoring as the singer's ability to entertain the crowd; the player could be cut off mid-song if the "crowd meter" dropped too low due to out-of-tune singing. This format became the new standard that subsequent karaoke games would imitate.

The following year, *SingStar* was released in Europe and Oceania, following the template laid down by *Karaoke Revolution*. They were joined in 2007 by *Rock Band*, and shortly afterwards by *Guitar Hero World Tour*, which expanded the karaoke format to include plastic guitars and drums that could be played alongside the vocals. The *Lips*, *Boogie* and *Sing It!* series also joined the karaoke stable at this time. Many of these series released multiple games per year across multiple gaming platforms, and every major game console featured at least one karaoke series.

In 2011, near the height of the karaoke boom, Fletcher and Light (2011) conducted two ethnographies of *SingStar* player groups. They describe *SingStar* as a "glue technology" (Fletcher and Light, 2011, p. 1) that facilitated social bonding and inter-generational connections within the groups. They also documented a variety of collaboration and co-play practices, wherein the microphone was shared between varying numbers of players, with secondary singers often acting as a social support to the primary player holding the microphone.

The karaoke craze lasted through two generations of consoles. From the launch of *Karaoke Revolution* in 2003 to the end of 2012, more than 120 different karaoke games were released (counting multi-platform releases as a single game). More than 70 of these were published in just three years, between 2008 and 2010. But as the market became oversaturated, consumer interest in the genre finally waned. In 2011 the rate of new releases began to wind down, and two attempts to revive the genre with new games in 2015 met with little success (David, 2016; Yin-Poole, 2016).

The Kinect Effect

By 2010, when the karaoke craze was at its peak, other forms of voice interaction gaming had fallen into a lull. The Nintendo DS was six years old and soon to be superseded, and had always been constrained by the limited sensitivity of its microphone. Voice-controlled strategy games had failed to gain traction after the limited success of *EndWar*, and the *SOCOM* series that pioneered voice control in the tactical shooter genre had stopped including it as a feature.

The situation was reversed in late 2010 when Microsoft released the Kinect, a multi-purpose sensor bar for the Xbox 360. The Kinect was primarily marketed as a motion sensing camera, which could track players' movements in three-dimensional space. However, it also included a microphone array and a speech recognition system. This provided a ready-made facility for game developers to add voice interaction to their games without needing in-house expertise in speech recognition. It also quickly established a substantial consumer market for voice interaction games, as within its first two months the Kinect became the fastest-selling consumer electronics device in history (Peckham, 2011).

This resulted in a new wave of voice interaction games, although it took some time to gather momentum as most early Kinect games focused on motion control rather than voice interaction. The launch title *Kinectimals*, a virtual pet game following the *Nintendogs* template, allows players to teach their animal tricks using voice commands such as "sit down" and "roll over". The following year, limited voice controls were included in *Dance Central 2*, *Kinect Sports: Season 2* and *Halo: Combat Evolved Anniversary*. In 2012, the sci-fi squad-based shooters *Binary Domain* and *Mass Effect 3* both allowed players to give verbal orders to their teammates in battle, and to converse with characters by reading out scripted lines of dialogue. The crest of the wave came in 2013, with the launch of the Xbox One console. For the first six months of its release, every Xbox One came bundled with a Kinect as a mandatory add-on (Turner, 2014), further expanding the potential user base for voice interaction. There followed a run of games that took advantage of voice input, in a diverse range of genres: action games such as *Ryse: Son of Rome*; family-friendly fare such as *Zoo Tycoon*; sports games such as *FIFA 14*; racing games such as *Forza Motorsport 5*; and role-playing games such as *Dragon Age: Inquisition*. More than 30 major releases to date have used the Kinect's voice recognition. The voice affordances vary considerably between these games, although it is notable that all of them make voice input an optional feature rather than a core element of the design – perhaps reflecting the fact that many Xbox One owners do not have a Kinect. In most of these cases, voice commands are simply an alternative means of activating actions that are already afforded by the primary control scheme; they form a redundant control layer rather than a novel implementation of voice-specific game mechanics.

Not all the Kinect games use spoken commands in this way, however. Some focus on creating the illusion that the characters in the gameworld are simply able to hear the sound of the player. The stealth horror game *Alien: Isolation* includes an optional "noise detection" mode, in which any moderately loud noise detected by the Kinect microphone alerts the game's fearsome antagonist (the titular alien) to the location of the player-character. Similarly, the stealth action game *Splinter Cell: Blacklist* allows the player to call out to in-game enemies, prompting them to investigate.[6] This feature was well received by players and reviewers alike for supporting a sense of connection with the gameworld (Carter, Allison, Downs, & Gibbs, 2015, p. 267).

Unlike in earlier eras, player feedback on the voice interaction features of Kinect games did not focus on the accuracy of speech recognition, which was generally considered to be acceptably reliable for many accents. Instead, voice interaction was critically appraised largely according to two considerations: its efficiency compared to regular controls, and its appropriateness within the context of the game's imaginary and overall design (Carter et al., 2015, p. 266). Voice commands that allowed players to role-play momentarily as their character were praised, whereas voice commands that called attention to the player's position outside the gameworld were criticised as feeling "unnatural" and "embarrassing" - an effect that Carter et al. termed "identity dissonance" (2015, p. 268).

Add-ons, Mods and Open Platforms

Most of the games described so far have been the products of well-established development and publishing studios. One reason for this is that the early game platforms on which voice

[6] A similar trick had been included ten years earlier in *Manhunt*, another stealth action game.

interaction was enabled were consoles, which had higher barriers to access and distribution than home computers. However, the greater openness and flexibility of the home computer as a platform has meant that voice interaction gaming has been available to those willing to seek out extra software or game mods.

The predecessors to today's more open approach to voice interaction games emerged in the 1980s, when speech recognition systems became available for home PCs that included voice control schemes for virtual flight simulators (Schoen, 1985). By the end the 1990s, such voice control software was marketed specifically for PC games, including Sontage Interactive's Game Commander and Microsoft Sidewinder Game Voice. These systems worked by converting spoken commands into user-specified keyboard or mouse inputs. Thus the voice control was a separate overlay rather than an integrated part of the game itself, although a few games came with pre-configured profiles for voice control software, including *Star Trek: Bridge Commander.*

Over time, as free speech recognition software became more accessible, some players began to take it upon themselves to add voice interaction to PC games by creating and distributing unofficial game modifications ("mods") that leveraged software from open source speech recognition systems such as CMU Sphinx. A notable example of player-led development of voice interaction is the role-playing game *The Elder Scrolls V: Skyrim*, by Bethesda Game Studios. Magical incantations are a central action in Skyrim's imaginary and game mechanics, but when it was first released in 2011, the game had no voice interaction features. Modders quickly responded by creating mods that enabled players to perform their incantations vocally;

one such mod had been downloaded by more than 49,000 different users as of June 2017.[7] In 2012, Bethesda responded by adding similar voice-command features in an official update to the Xbox 360 version of the game.

In the four years since voice commands were brought to *Skyrim*, there has been a proliferation of smaller-scale, smaller-budget voice interaction games produced by independent developers. These have appeared for both PCs (*Bot Colony*, *There Came an Echo*, *In Verbis Virtus*, *Plan Be*) and mobile phones (*The Howler*, *Ah! Bird*, *Mayday! Deep Space*), the latter having the advantage that microphones are a standard feature of the platform. Notably, all of these small-scale independently-developed games have made voice interaction a central focus of their player experience; several are unplayable without a microphone. As a result, their design engages more closely with the affordances of voice than most larger-studio productions, for which voice input is merely a bonus. For example, both *Plan Be* and *Mayday! Deep Space* are stealth games that provide the player with a "radio" connection to an in-game character, whom they must guide through a dangerous facility with verbal directions while looking at an abstracted overhead map (Figure 5). Through the voice of the character "on the ground", these two games evoke a richer sensory environment than they depict visually, creating a heightened tension and allowing the player's imagination to fill in the scene (Shimomura, 2015). Reviews attributed a heightened experience of social involvement to the spoken interaction format:

> This is the rarely explored power of voice: speaking to a character with your own actual

[7] 'ThuuMic – Use your mic for dragon shouts' by DeadlyAzuril and PsychoHampster, retrieved 8 June 2017, from http://www.nexusmods.com/skyrim/mods/5626/

> voice connects you to them, forges a bond. (Shimomura, 2015)

Discussion

Voice interaction has had a place in digital games, for research and entertainment, since 1973. But its path towards acceptance and popularity has not been smooth. Hardware and software platforms have acted as constraints on the design space of voice interaction, leading to a tidal pattern in which voice interaction games have repeatedly risen on a wave of novelty with a new game platform, only to fall away as its limitations are discovered. We have shown that the development of voice interaction game design has been responsive to the platforms on which digital games are played, but that this influence is not deterministic, as developers in different countries have taken distinctive approaches to their use of the technology.

Platform Reactivity

Although voice interaction gaming had something of a false start in the 1980s, achieving only limited success with the severely limited voice input system on the Famicom, it is notable how many different genres of voice interaction games were present in those early years. From karaoke singing to "bargaining" with merchants to blowing air and directing a sports team, the games of that era foreshadowed many of the genres that would emerge on more capable platforms. Nevertheless, the technology could not support the ambitions of game makers or the demands of players. The long winter that followed for voice interaction games lasted until game platforms once again found a reason to include microphones, with the advent of online gaming. That latent consumer interest in voice interaction remained throughout this period is suggested by the stratospheric success of the Nintendo DS, which was the first platform to foreground

voice interaction as a central game mechanic in its most popular titles. We do not mean to imply that voice interaction was the primary reason for the success of the DS, but its prominence in the design and marketing of bestselling games such as *Brain Age* and *Nintendogs* suggests that it was a contributing factor. On other platforms around the same time, the digital companion games and tactical combat games that adopted voice interaction attracted a smaller audience, reflecting the lower market penetration of headsets and other microphones on these consoles.

What this history has clearly shown is that each boom in voice interaction has proceeded from a new platform development, beginning with the Famicom microphone and continuing most recently with smartphone gaming. Voice interaction game design has been both driven and constrained by the affordances of the technology platforms on which digital games have been built at each point in time. While examples of voice interaction exist in almost every era of digital games, the great majority have closely followed the release of a platform that afforded new voice input or speech processing capabilities. This highlights the importance of studying platforms as context for understanding the games that are played upon them, as exemplified in the interlinked fields of platform studies (Bogost & Montfort, 2009) and media archaeology (Apperley & Parikka, 2015; Parikka & Suominen, 2006). The patterns we have observed in this history illustrate the influence dynamic that Bogost and Montfort ascribe to platforms, wherein the technology is not deterministic but does "encourage and discourage different sorts of expressive new media work" (2009, p. 5).

Karaoke games may be considered as an exception to the platform-driven rule, as games in this genre have often been sold with the microphones that enable their use, rather than taking

advantage of pre-existing microphones that come with the console on which they are played. This can be explained by the extra functions that microphones play in karaoke games, as a social prop and a cultural referent (Fletcher & Light, 2011, p. 11) as well as a controller. Whereas in most digital games the imagined action is situated in the virtual space and oriented towards virtual characters, in karaoke games the imagined action is situated in the room and (often) oriented towards a real audience. The karaoke genre has traditionally used large, stage-style microphones rather than discreet headsets or controller microphones; these provide both a clear visual cue for the audience to see the player in the context of a stage performance rather than an ordinary living-room scene, and a focal point away from the screen around which multiple singers can orient themselves (Fletcher & Light, 2011, p. 10). Thus, the voice input hardware is an even more essential feature of karaoke games than many other voice interaction games, for which the physical format of the microphone is less relevant.

Voice Interaction Game Cultures

To say that technology and platforms have driven the development of voice interaction games is not to say that they have fully determined how these game experiences have been designed. As this history has described, voice interaction games produced in different regions show signs of being the product of different ways of thinking about how players might communicate with virtual characters. This shows that cultural, individual and economic factors still play a large role in what kinds of experience we see as possible and desirable for voice interaction games.

The clearest stylistic differences are those that emerged between the voice interaction games made in Japan and those made in North America and Europe. This is particularly apparent in the

way that each of these regions typically approached the player's relationship with the virtual characters in the games. Japanese developers have tended to create conversational interactions in which the player engages in two-way communication with a singular character who is presented as an autonomous being. This style is exemplified by *Hey You, Pikachu!* and *Seaman*, and might be called "digital companion" or "virtual pet" games. While the virtual character is clearly responsive to the player's inputs rather than fully independent, they are presented by the game as being in a reciprocal relationship with the player, in which relationship building is at least partly the goal. In contrast, Western developers have typically created militaristic strategy games in which the player uses their voice to issue commands to multiple units, in an authoritative, top-down communication style that suits the hierarchical structure of the player-character relationship; the inner thoughts of the virtual characters are of minimal relevance. *Tom Clancy's EndWar* and *Command: Aces of the Deep* are archetypal examples. While not every voice interaction game from a Japanese, North American or European developer fits these profiles, there is a strong tendency towards these styles of voice interaction among most of the games from each region.

Interestingly, the stylistic differences we have noted alternately support and contradict the argument made by Nass and Brave (2005) that all speech interaction automatically evokes an imagined social relationship with the technology. Nass and Brave report on a number of experiments that suggest people's responses to computers accord with their responses to other humans in equivalent interactions, and that this effect is particularly pronounced for voice interactions. This effect might be expected to be even more pronounced in digital games, as virtual characters are such a central feature of the medium. Indeed, the Japanese style of two-

way, relatively equal communication humanises (or anthropomorphises) the virtual characters and suggests a relationship that very much treats the game character as a true conversation partner with a mind of its own. Conversely, the Western style of top-down control implies a much less humanised view of the virtual characters, and a greater sense that the player is simply moving tokens on a game-board.

The peculiarities of Western and Japanese game design have long been identified by contrast with one another, although it would be reductive to suggest that there is one monolithic culture or approach to game design on either side (Consalvo, 2016, pp. 4–5). The differences we have observed may be due to broad cultural factors, or they may be the result of individual preferences; the number of people making voice interaction games has always been small enough that a single developer could have a substantial influence on the style of games produced in their region. What these differences clearly demonstrate is that the content and style of voice interaction games have not simply been determined by the affordances of the platforms on which they are built, although these platforms have certainly shaped what is possible.

The tidal pattern of voice interaction gaming that this history has documented may now be coming to an end, as microphones and voice interfaces become an increasingly standard component of gaming devices. There appears to be little prospect of a return to a situation like the winter of voice interaction games in the 1990s, when common gaming platforms simply did not feature voice interfaces due to hardware constraints. At the same time, the barriers to developing and publishing a voice interaction game have been lowered by open-source speech processing systems and digital distribution channels, which has enabled the proliferation of

smaller-scale games by independent developers. Without the platform-driven cycle of novelty to rely on, designers will need to understand how past attempts at voice interaction in games have succeeded or failed in engaging users. The history outlined in this article has shown that videogames have a long record of experimentation and creativity with the format, much of it let down by the limitations of the technology of the day. Recognising what kinds of interactions are desirable, and how different configurations of the player-character relationship create meaning for players, will benefit both game designers and developers of voice interfaces for other contexts.

References

150 Best Games of All Time. (1996, November). Computer Gaming World, (148), 63–80.

Altice, N. (2015). *I Am Error: The Nintendo Family Computer / Entertainment System Platform*. Cambridge, MA: MIT Press.

Apperley, T., & Parikka, J. (2015). Platform Studies' Epistemic Threshold. *Games and Culture*, 1555412015616509. https://doi.org/10.1177/1555412015616509

Aylett, M. P., Kristensson, P. O., Whittaker, S., & Vazquez-Alvarez, Y. (2014). None of a CHInd: Relationship Counselling for HCI and Speech Technology. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems* (pp. 749–760). New York: ACM. https://doi.org/10.1145/2559206.2578868

Baker, J., Deng, L., Glass, J., Khudanpur, S., Lee, C.-H., Morgan, N., & O'Shaughnessy, D. (2009). Developments and directions in speech recognition and understanding, Part 1 [DSP Education]. *IEEE Signal Processing Magazine*, *26*(3), 75–80. https://doi.org/10.1109/MSP.2009.932166

Berger, C. (1916, December 19). Sound-operated circuit-controller. Retrieved 15 August 2017, from https://worldwide.espacenet.com/publicationDetails/biblio?II=0&ND=3&adjacent=true&locale=en_EP&FT=D&date=19161219&CC=US&NR=1209636A&KC=A

Berger, C. (1917, October 16). Automatically-operating toy or the like. Retrieved 15 August 2017, from https://worldwide.espacenet.com/publicationDetails/biblio?II=0&ND=3&adjacent=true&locale=en_EP&FT=D&date=19171016&CC=US&NR=1243380A&KC=A

Berger, C. (1918, September 24). Sound-operated toy or instrument. Retrieved 15 August 2017, from http://worldwide.espacenet.com/publicationDetails/biblio;jsessionid=cmLm1-lKKSLAdeqyB-BeTcCz.espacenet_levelx_prod_2?FT=D&date=19180924&DB=&locale=&CC=US&NR=1279831A&KC=A&ND=1

Bogost, I., & Montfort, N. (2009). Platform studies: Frequently questioned answers. *Digital Arts and Culture*.

Carter, M., Allison, F., Downs, J., & Gibbs, M. (2015). Player Identity Dissonance and Voice Interaction in Games. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play* (pp. 265–269). New York: ACM. https://doi.org/10.1145/2793107.2793144

Cohen, M. H., Giangola, J. P., & Balogh, J. (2004). *Voice User Interface Design*. Boston, MA, USA: Addison-Wesley Professional.

Consalvo, M. (2016). *Atari to Zelda: Japan's Videogames in Global Contexts*. Cambridge, MA: MIT Press.

Dark Watcher, & 98PaceCar. (n.d.). RDI Halcyon. Retrieved 29 November 2016, from http://www.videogameconsolelibrary.com/pg80-rdi.htm

David, E. (2016, February 12). Was the world not ready for Rock Band 4 and Guitar Hero Live? Retrieved 29 November 2016, from http://siliconangle.com/blog/2016/02/12/was-the-world-not-ready-for-rock-band-4-and-guitar-hero-live/

Dow, S., Mehta, M., Harmon, E., MacIntyre, B., & Mateas, M. (2007). Presence and Engagement in an Interactive Drama. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1475–1484). New York: ACM. https://doi.org/10.1145/1240624.1240847

Fairbanks, A. (Ed.). (1931). *Philostratus, Imagines. Callistratus, Descriptions.* London: William Heinemann Ltd.

Karaoke Studio. (n.d.). Retrieved 7 June 2017, from http://famicomworld.com/system/other/karaoke-studio/

Fletcher, G., & Light, B. (2011). Interpreting digital gaming practices: SingStar as a technology of work. Presented at the European Conference on Information Systems, Helsinki, Finland. Retrieved 15 August 2017, from http://usir.salford.ac.uk/17245/

Furui, S. (2005). 50 years of progress in speech and speaker recognition. *SPECOM 2005, Patras*, 1–9.

Gernsback, H. (1916, June). An electric pup that heeds your call. *The Electrical Experimenter*, *4*(2), 104.

Harada, S., Wobbrock, J. O., & Landay, J. A. (2011). Voice Games: Investigation into the Use of Non-speech Voice Input for Making Computer Games More Accessible. In *Proceedings of the 13th IFIP TC 13 International Conference on Human-computer Interaction -*

Volume Part I (pp. 11–29). Berlin, Heidelberg: Springer-Verlag.

Herman, L. (1997). *Phoenix: The Fall & Rise of Videogames* (2nd ed.). Rolenta Press.

Huang, X., Baker, J., & Reddy, R. (2014). A Historical Perspective of Speech Recognition. *Communications of the ACM*, *57*(1), 94–103. https://doi.org/10.1145/2500887

Igarashi, T., & Hughes, J. F. (2001). Voice As Sound: Using Non-verbal Voice Input for Interactive Control. In *Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology* (pp. 155–156). New York: ACM. https://doi.org/10.1145/502348.502372

Kinder, J., & Hallock, D. (n.d.). Halcyon Interactive Laserdisc System. Retrieved 13 December 2016, from http://www.dragons-lair-project.com/community/related/homesystems/halcyon/

Lasers and Computers. (1985, January 14). *Computer Chronicles*. USA: PBS.

Lea, W. A., & Shoup, J. E. (1979). *Review of the ARPA SUR Project and Survey of Current Technology in Speech Understanding.* (No. N00014-77-NaN-570) (p. 162). Speech Communications Research Lab, Los Angeles CA: Office of Naval Research.

Lee, K.-F. (1988). Word Recognition in Large Vocabularies: On large-vocabulary speaker-independent continuous speech recognition. *Speech Communication*, *7*(4), 375–379. https://doi.org/10.1016/0167-6393(88)90053-2

Luger, E., & Sellen, A. (2016). 'Like Having a Really Bad PA': The Gulf Between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 5286–5297). New York: ACM. https://doi.org/10.1145/2858036.2858288

Lynch, K. (2013, November 18). PlayStation 4: Can Sony's new console live up to its

predecessor's record-breaking legacy? Retrieved 29 November 2016, from http://www.guinnessworldrecords.com/news/2013/11/playstation-4-can-sony%E2%80%99s-new-console-live-up-to-its-predecessor%E2%80%99s-record-breaking-legacy-47090

Mace, S. (1983, November 21). TI retires from home-computer market. *InfoWorld*, *5*(47), 22–27.

Munteanu, C., Jones, M., Oviatt, S., Brewster, S., Penn, G., Whittaker, S., … Nanavati, A. (2013). We Need to Talk: HCI and the Delicate Topic of Spoken Language Interaction. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (pp. 2459–2464). New York: ACM. https://doi.org/10.1145/2468356.2468803

Mustaquim, M. M. (2013). Automatic speech recognition: an approach for designing inclusive games. *Multimedia Tools and Applications*, *66*(1), 131–146. https://doi.org/10.1007/s11042-011-0918-7

Nanjo, H., Mikami, H., Kunimatsu, S., Kawano, H., & Nishiura, T. (2009). A Fundamental Study of Novel Speech Interface for Computer Games. In *ISCE: 2009 IEEE 13th International Symposium on Consumer Electronics, Vols 1 and 2* (pp. 291–293). New York: IEEE.

Nass, C., & Brave, S. (2005). *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*. Cambridge, MA: MIT Press.

Parikka, J., & Suominen, J. (2006). Victorian Snakes? Towards A Cultural History of Mobile Games and the Experience of Movement. *Game Studies*, *6*(1). Retrieved 15 August 2017, from http://gamestudies.org/0601/articles/parikka_suominen

Peckham, M. (2011, March 9). Kinect Breaks Guinness Record, Sells 10 Million Systems, Tops iPhone and iPad. Retrieved 29 November 2016, from

http://www.pcworld.com/article/221738/kinect_breaks_guinness_record_sells_10_millions_systems.html

Provo, F. (2000, November 3). Hey You, Pikachu! Review. Retrieved 18 May 2016, from http://www.gamespot.com/reviews/hey-you-pikachu-review/1900-2650135/

Rabiner, L., & Juang, B.-H. (2008). Historical Perspective of the Field of ASR/NLU. In J. Benesty, M. M. Sondhi, & Y. Huang (Eds.), *Springer Handbook of Speech Processing* (pp. 521–538). Berlin, Heidelberg: Springer.

Reddy, D. R., Erman, L., & Neely, R. (1973). A model and a system for machine recognition of speech. *IEEE Transactions on Audio and Electroacoustics*, *21*(3), 229–238. https://doi.org/10.1109/TAU.1973.1162456

Reed, K. (2008, November 6). Tom Clancy's EndWar Review. Retrieved 29 November 2016, from http://www.eurogamer.net/articles/tom-clancys-endwar-review

Schoen, J. (1985, March 5). When You Talk, Your PC Listens. *PC Mag*, *4*(5), 122–132.

Shimomura, D. (2015, January 21). Mayday! Deep Space relies on the panic in your mind. Retrieved 8 June 2017, from https://killscreen.com/articles/mayday-deep-space-relies-panic-your-mind/

Sporka, A. J., Kurniawan, S. H., Mahmud, M., & Slavík, P. (2006). Non-speech input and speech recognition for real-time control of computer games (Vol. 2006, pp. 213–220). Presented at the Eighth International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS 2006. https://doi.org/10.1145/1168987.1169023

Stevens, N. (1999, October). Command Aces of the Deep. Retrieved 29 November 2016, from http://www.subsim.com/ssr/commandaces.html

Suominen, J. (2016). How to Present the History of Digital Games: Enthusiast, Emancipatory,

Genealogical, and Pathological Approaches. *Games and Culture*. https://doi.org/10.1177/1555412016653341

Toys that obey your voice. (1916, December). Popular Science Monthly, 89, 718–719.

Turner, A. (2014, May 14). Microsoft backflips on Kinect for Xbox One. Retrieved 20 May 2016, from http://www.smh.com.au/digital-life/computers/gadgets-on-the-go/microsoft-backflips-on-kinect-for-xbox-one-20140514-zrcc7.html

Wadley, G., Carter, M., & Gibbs, M. (2015). Voice in Virtual Worlds: The Design, Use, and Influence of Voice Chat in Online Play. *Human–Computer Interaction*, *30*(3–4), 336–365. https://doi.org/10.1080/07370024.2014.987346

Warren, T. (2014, April 2). The story of Cortana, Microsoft's Siri killer. *The Verge*. Retrieved 15 August 2017, from http://www.theverge.com/2014/4/2/5570866/cortana-windows-phone-8-1-digital-assistant

Wolf, M. J. P. (2008). *The Video Game Explosion: A History from PONG to Playstation and Beyond*. Santa Barbara, CA: ABC-CLIO.

Wolf, M. J. P. (2012). *Encyclopedia of Video Games: The Culture, Technology and Art of Gaming* (Vol. 2). Santa Barbara, CA: Greenwood.

Yin-Poole, W. (2016, February 12). Guitar Hero Live failed to do the business, too. Retrieved 29 November 2016, from http://www.eurogamer.net/articles/2016-02-12-guitar-hero-live-failed-to-do-the-business-too

Acknowledgements

This work was supported by the Microsoft Research Centre for Social NUI and the Australian Government Research Training Program.

Author Biographies

Fraser Allison is a PhD candidate at the Microsoft Research Centre for Social NUI at the University of Melbourne. His doctoral research examines the use of conversational interfaces to interact with virtual characters.

Marcus Carter is a lecturer in digital cultures at the University of Sydney. Please refer to his website: http://marcuscarter.com/

Martin Gibbs is an associate professor in the School of Computing and Information Systems at the University of Melbourne. His current teaching and research interests lie at the intersection of science, technology studies (STS) and human–computer interaction, and are focused on the sociable use of interactive technologies with a particular interest in the study of digital and analogue games.

Figures

C. BERGER.

SOUND OPERATED CIRCUIT CONTROLLER.

APPLICATION FILED MAY 19, 1916.

1,209,636.

Patented Dec. 19, 1916.

2 SHEETS—SHEET 2.

Fig. 4.

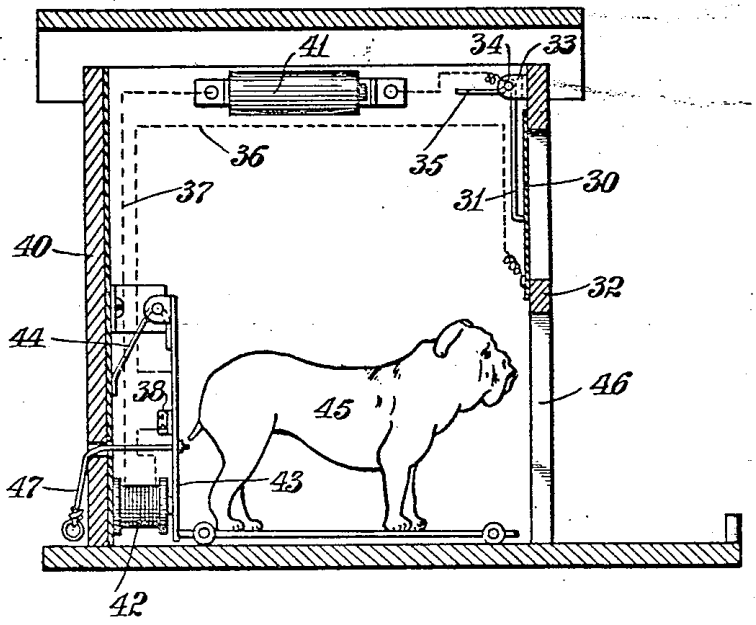

Fig. 5.

Fig. 6.

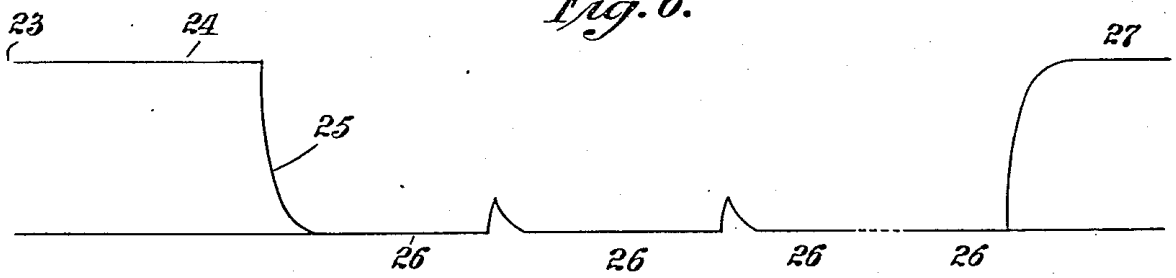

Figure 1. Patent illustrations for Radio Rex (Berger, 1916, p. 2).

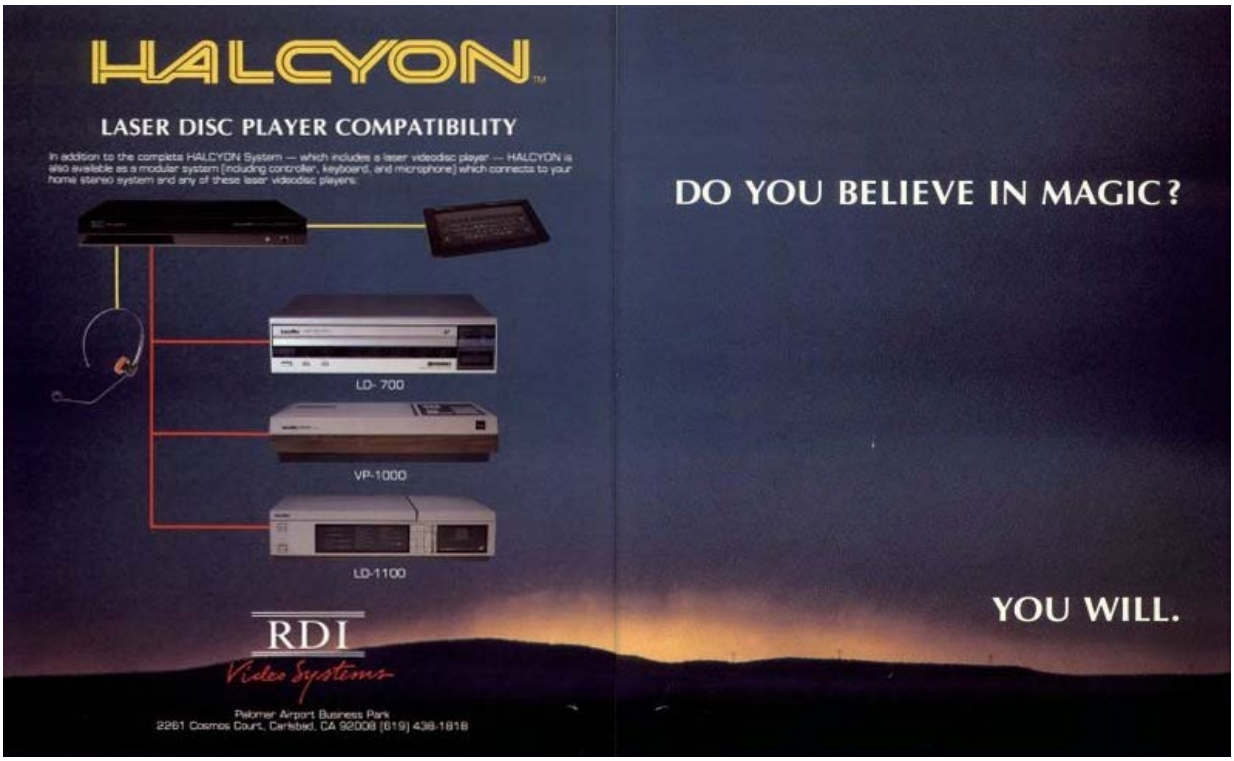

Figure 2. Flyer for the Halcyon (Kinder & Hallock, n.d.).

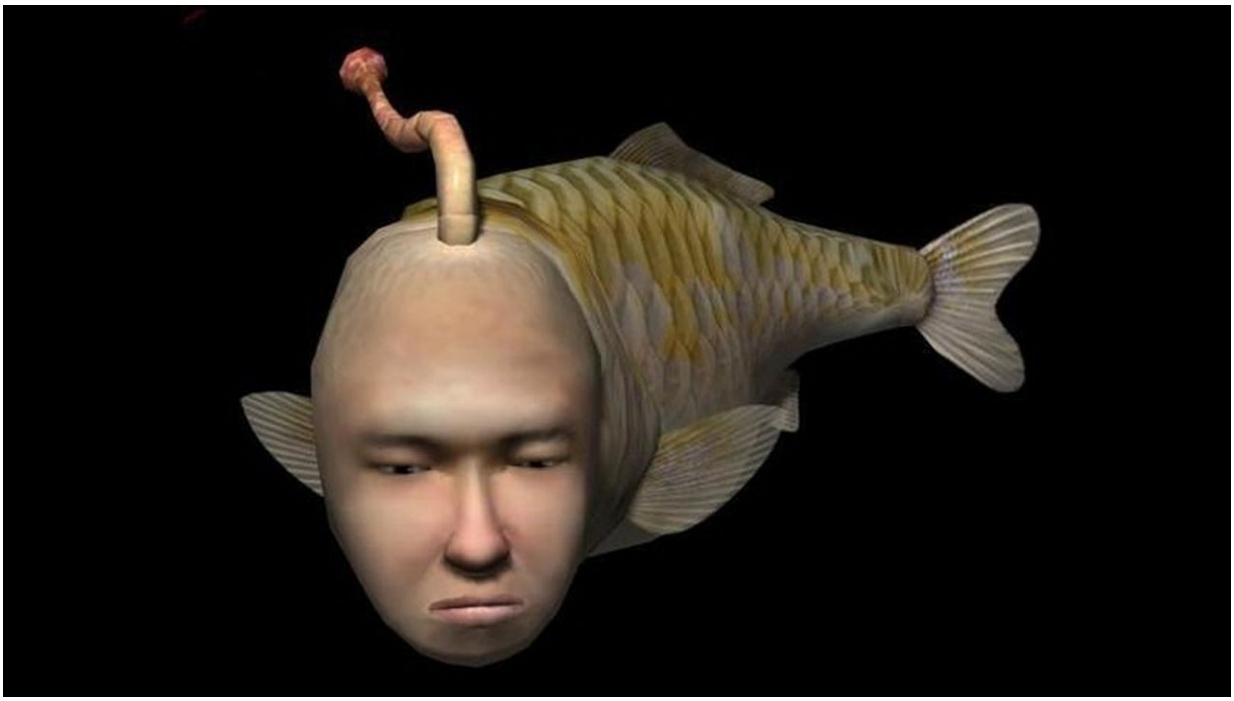

Figure 3. The main character in *Seaman*, a fish with a human face.

Figure 4. Piano roll-style pitch indicator in *Karaoke Revolution*.

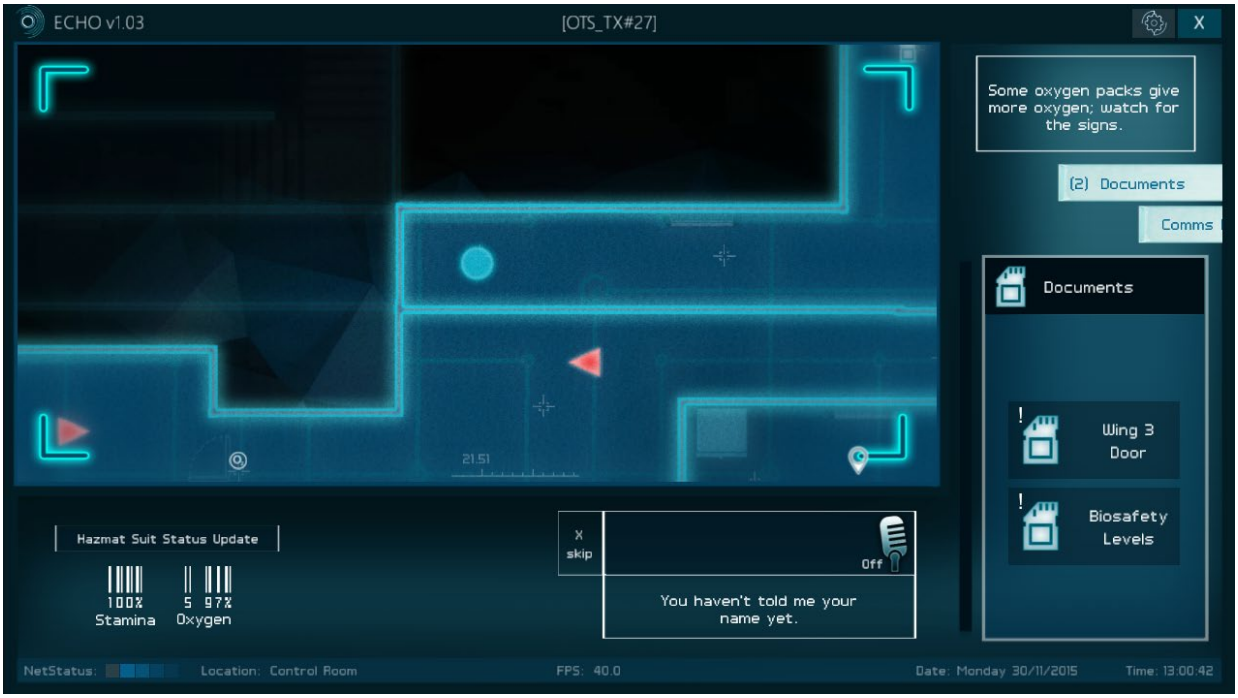

Figure 5. Overhead map view in *Plan Be*.

University Library

A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Allison, F;Carter, M;Gibbs, M

Title:
Word Play: A History of Voice Interaction in Digital Games

Date:
2020

Citation:
Allison, F., Carter, M. & Gibbs, M. (2020). Word Play: A History of Voice Interaction in Digital Games. Games and Culture, 15 (2), pp.91-113. https://doi.org/10.1177/1555412017746305.

Persistent Link:
http://hdl.handle.net/11343/282434