Harvester of Desires: Gaming Amazon Echo through John Cayley's *The Listeners*

Laura Okkema
University of Central Florida
4000 Central Florida Boulevard
+1 906 231 2525
laura.okkema@gmail.com

ABSTRACT
Over the past two years, smart speakers such as Amazon Echo have become popular entertainment technologies and, increasingly, game platforms in households across the globe. These systems are controlled through voice-interactive Artificial Intelligences such as Amazon's Alexa. The present work seeks to open a conversation about voice-interactive games on smart speaker systems in game studies. While these platforms open exciting new creative spaces for gamers and game developers alike, they also raise ethical concerns: Smart speakers are powerful twenty-first century surveillance capable of interpreting, recording and synthesizing human speech. Through the lens of a case study on John Cayley's ludic Alexa skill *The Listeners*, this paper interrogates how Amazon Echo's technological affordances enable new forms of surveillance while also giving rise to a new poetics of voice interaction. Illuminating aesthetic and ethical dimensions can help scholars in game studies assess the risks and perks of this new ludic platform.

Keywords
Alexa, ethics, Amazon Echo, poetics, platform studies, transactive synthetic voice, voice-interactive games, surveillance

INTRODUCTION
Voice interaction in games has received little attention in game studies during the past decade, but recent technological advances in transactive synthetic language and resulting changes on the market for entertainment technologies have given significance to the subject (Allison et al. 2017, 2). Smart speaker systems have become highly popular since the launch of Amazon Echo in 2015, so much that these voice-controlled audio technologies are now found in 16 % of American homes (NPR & Edison Research 2017). The voice interface is accessed via an Artificial Intelligence, such as Amazon's Alexa, Google Assistant, and Apple's Siri, all of which are capable of processing natural human language, meaning users can initiate the various functions of their smart speaker system by talking to the device (Cayley 2016).

Smart speaker owners frequently use their system to listen to music, order food, obtain directions, or control devices around the house including door locks, lights, or heating; all the while, the device will learn about its owner's habits, needs, and preferences, and advertise products and businesses whenever possible (Dale 2017, 970; NPR & Edison Research 2017). Playful interactions with the AI are a key part of the systems' appeal for consumers, especially to entertain family and friends. Popular activities include asking

trivia questions, having the AI tell jokes, and playing voice-interactive games (NPR & Edison Research 2017). Users can create their own applications, so called *Alexa* skills, for Amazon Echo, which opens rich possibilities for not only playing, but also making one's own games. With these features, smart speaker systems afford new spaces for creative ludic expression, as evidenced by the over 6000 registered *Alexa* skills that fall under the heading of "Entertainment, Games and Trivia."

Yet as a new poetics of voice interactive games emerges from smart speaker platforms, so do new forms of invading user privacy. Amazon Echo's Terms and Conditions state that once activated with the trigger word, the device processes and records everything that is said in its presence, and stores the data on the registered user's account as well as in Amazon's databases (Dale 2017, 970). A particularly unsettling thought is the notion that guests who play a game with their host's *Alexa* may not know they are being recorded by a corporation (such as Amazon) *and* their host. What if these guests proceed to discuss personal matters while the smart speaker system is still active? There is no legal obligation for Alexa owners to inform their guests that the smart speaker is listening. Due to the relative novelty of synthetic transactive voice technologies, situations such as these remain unregulated by the law (Cayley 2016).

Combining a platform studies approach to the Amazon Echo with a close reading of John Cayley's ludic Alexa skill *The Listeners,* this paper interrogates how Amazon Echo's technological affordances enable new forms of surveillance while also giving rise to a new poetics of voice interaction. Cayley's game both instantiates and comments on the problematic nature of smart speaker systems by showing how the entertainment industry gathers information about users' desires and emotions for profit. *The Listeners* thus serves as a lens to understand how playful applications on new platforms like Amazon Echo are intertwined with corporate interest and systemic surveillance. The following sections serve to embed this discussion within current discourses in game studies, specifically with an eye to voice interactive games and platform studies. After a brief overview of my methods, I discuss Amazon Echo's technological affordances and the case study, John Cayley's *The Listeners*.

AN OVERVIEW OF VOICE INTERACTION IN GAMES

The availability of games which use voice-based interfaces has largely depended upon developments in speech recognition and the accessibility of hardware such as microphones and speakers. In their study of hardware evolution in video games, Jörnmark et al. showed that the game industry's economic success has been fueled by innovation and incorporation of new hardware: "[f]or every gaming platform, games have developed new expressions and forms" (2005, 3). Voice-based interfaces are no exception; the emergence of voice-interactive features in games is closely intertwined with technological innovation in speech recognition and the availability of suitable platforms.

Attempts to build machines that can 'listen' go back to the mid-eighteenth century, but successful implementation did not happen until the 1960s (Juang & Rabiner 2004, 4). Early speech recognition software could tackle small vocabularies of 10-100 isolated words and found application in manufacturing companies for quality control. Technological breakthroughs in the 1980s made possible the recognition of continuous speech and vocabularies of 1000 words and more (Juang & Rabiner 2004, 20). Around this time, voice recognition features were added to existing gaming platforms for the first time, among them the "MBX" module for Texas Instruments' TI-99/4A computer, which featured seven

speech interactive games (Allison et al. 2017, 6). However, none of these early attempts at incorporating voice interaction in gaming consoles were commercially successful.

The first profitable game platform with voice-interactive features was Nintendo's Family Computer (Famicom). The Famicom had a simple microphone built into one of its controllers; it was ignored by game developers until 1985. The popularity of the platform instead lay in a combination of "its price, hardware capacity and innovative game library" (Jörnmark et al. 2005, 6). However, Allison et al. note in their genealogy of voice interaction in games that once developers began to experiment with voice interaction, the Famicom offered a surprising variety of uses for voice-based input despite the hardware's limitations, "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*)" (Allison et al. 2017, 6). While voice interaction may have been among the Famicom's appealing features, it remained on the sideline as a form of input, and in the following decades continued to serve as an addition or alternative to other forms of existing controls.

Vocal interfaces in games have primarily supplemented visual and kinesthetic forms of input. Allison et al. identify two early trends in the implementation of voice-based features. In Japanese games, voice interaction has traditionally focused on relationship-building. Driven largely by Nintendo's experiments with different forms of voice input, this type of interactivity entered the console market in the late 1990s with games like *Hey you, Pikachu!* (1998), in which players could use voice interaction to befriend and converse with a friendly Pokémon (Allison 2017, 8-9). By contrast, voice interaction on the Western game market has historically taken the form of command-and-control communication. An early example is *Command: Aces of the Deep* (1995), a World War II combat game which bears three features commonly found in Western voice interactive games: For one, the game "puts the player in the role of a military authority figure in a combat situation, who barks orders at a team of underlings;" secondly, the player's team consists of passive characters who lack complexity and are built to bend to the players will (Allison 2017, 10). Finally, voice interaction is not essential to playing the game. The commands can alternatively be given via mouse or keyboard interactions. Similarly, games for Microsoft's Kinect for the Xbox 360 (2010) incorporated some voice interaction as an alternative interface to existing modes of control.

Games specifically developed for voice control input are primarily found in the Karaoke genre, which became popular in the Western world when console-based games like *Sing Star* (2004) and *Rock Band* (2007) entered the market (Harada et al. 2011, 14). Other examples of games which place voice-control at the center include Flash-based side scroll shooter games like *PAH!* or *shout and dodge,* which "continuously map the volume of the microphone input to the vertical position of a spaceship" (Harada et al. 2011, 15). While specifically built for voice input, these games use little to no verbal recognition for gameplay, but rely on pitch, volume, and simple plosive sounds for control. Moreover, all previously mentioned games had a visual and tactile interface in addition to voice control.

Contrary to technologies which use voice interaction alongside other forms of input, smart speaker systems like the Amazon Echo rely on voice as the primary interface (Porcheron 2018). The voice-only interface and the device's ability to constantly listen set smart speaker systems like Amazon Echo apart from earlier technologies which incorporate voice interaction. These features are part of the device's affordances, the capabilities that constrain and enable what developers and gamers can do with the device. Salter and Murray state that affordances are more than technological properties; rather, "they become

suggestions and frameworks for the works built upon them." (Salter and Murray 2014, 2). In other words, developers' and gamers' ideas and creative engagement with a given device are informed by the technology's specific properties.

METHODOLOGY

The present work uses a case study, John Cayley's The Listeners, to address some of the complex entanglements of technology, poetics, ethics, economics, and politics in games. A cross-disciplinary approach which borrows from game studies as well as literary and cultural studies is therefore appropriate to this inquiry. To account for the technological as well as the cultural and political contexts of the Amazon Echo, I engage a platform studies perspective alongside a critical close reading. However, this study does not aim for a comprehensive platform study of the Amazon Echo; rather, it borrows from the field to frame the Echo as a platform. In doing so, I emphasize two things: one, while the present study focuses on a single game, the argument I put forth has much wider implications; two, future research needs to attend to smart speaker platforms in depth and broaden the scope of cases to include other games and applications. The following section situates the Amazon Echo within the discourse of platform studies. With this discussion, I emphasize that to analyze the Amazon Echo and its applications ideologically, culturally, and politically, we must first obtain a thorough understanding of its technological functions.

To analyze *The Listeners*, I recorded over an hour of transactions between myself and the game, covering all the interactions which presented themselves during one session. I transcribed the recording in a word processing document. Much of the game consists of spoken text, specifically monologues given by *The Listeners;* for this reason, and because of the John Cayley's background in poetry, a close reading was chosen as the appropriate method for analysis. Combining the reading of John Cayley's *The Listeners* with a thorough examination of the Echo as a platform allows for an in-depth interpretation of the emerging poetics of voice interactive games vis-à-vis ethical concerns about surveillance.

UNDERSTANDING AMAZON ECHO AS A PLATFORM

Defining the term 'platform' is challenging because the notion itself is highly abstract. According to Montfort and Bogost, who introduced the term 'platform studies' in their book *Racing the Beam: The Atari Video Computer System* (2009), a platform is "whatever the programmer takes for granted when developing, and whatever, from another side, the user is required to have working in order to use particular software" (2). This means that platforms can take different forms; gaming consoles, software, or an operating system could all fall under the term's umbrella. The defining characteristic is that the platform constitutes a necessary basis on both the programmer's and the user's end: it is the ground upon which everything else is built. Montfort and Bogost situate platforms at the deep end of what they call the "five levels of digital media." The five-level model helps identify the focal point of creative works in digital media; it includes reception and operation, interface, form and function, code, and platform. In relation to the other levels, the platform is the abstract entity that lies underneath the more concrete level of code. The level-model reinforces the notion that the platform constitutes a basis or foundation for all other aspects and components of any given digital medium.

Every platform, be it an Arcade machine, a Nintendo 64 console, or an Amazon Echo, is developed, distributed, and used in cultural, social, and economic contexts. In their study of the Atari VCS, Montfort and Bogost demonstrate the importance of acknowledging how technologies are embedded within culture and history. As they chronologically discuss six VCS cartridges in detail, they not only address how the levels of code and hardware work

together, but also account for aspects such as the company culture at Atari, the economic motivations of developers, and gamer culture in the 1980s. Their work sketches the history of the platform from its beginnings in the late 1970s to its discontinuation in the early 1980s, and concludes with a reflection on the Atari's status as a venerable technological relic and its influence on gamers, game design, developers, programmers, and the culture of gaming as a whole. The in-depth analysis of the Atari VCS demonstrates the importance of context in the analysis of any digital medium: "A computational platform is not an alien machine, but a cultural artifact that is shaped by values and forces and which expresses views about the world, ranging from 'games are typically played by two players who may be of different ages and skill levels' to 'the wireless service provider, not the owner of the phone, determines what programs may be run'" (Montfort and Bogost 2009, 148). The value of the platform studies approach to the present study lies, then, in its ability to help identify how a technology's affordances shape, and are shaped by, the beliefs and values held by developers, gamers, and society at large.

As the first in a series of publications on platform studies, Montfort and Bogost's work constitutes an early blueprint for how future scholarship can engage the platform studies approach. Since 2009, the field of platform studies has diversified and gained momentum in digital media scholarship. Scholarship in platform studies has covered Nintendo Wii (Jones and Thiruvathukal 2012), the Commodore Amiga (Maher 2012), Flash (Salter and Murray 2014), the Nintendo and the Supernintendo Entertainment System (Altice 2015; Arsenault 2017), the BBC Micro (Gazzard 2016), and Minitel (Mailland and Driscoll 2017). Friedhoff (2013) examines the interactive storytelling software Twine from a platform studies approach. She demonstrates how Twine's technological affordances, its openness and accessibility to novice programmers, make the platform suitable for independent developers who want to address social justice concerns in their games. Allison et al.'s (2017) previously discussed study offers a history of voice-interactive platforms, but focuses on developments prior to smart speaker systems such as Amazon Echo.

Amazon Echo is an audio technology controlled via Alexa, an Artificial Intelligence capable of automatic voice recognition and synthesis. The "body" of the Amazon Echo consists of a loudspeaker and seven microphones. The name Alexa functions as a trigger word: if a user says "Alexa" loudly and clearly, the system is activated. Once the system is active, users can interact with it. Everything the microphones pick up is streamed to the cloud-based Alexa Voice Services. Automatic speech recognition software converts the captured audio into text, then Natural Language Understanding software breaks the text into chunks which Alexa can act on (Straczek et al. 2017). Once the speech input is processed in this manner, the software can complete the request. Finally, Alexa sends a directive to the Amazon Echo to output the generated file.

Part of the speech recognition process involves capturing and recording what is being said. These recorded transactions are not deleted; rather, everything Alexa hears "is fed back to the interlocutor who owns the account to which she [Alexa] is registered;" moreover, every interaction is registered and stored by Amazon Voice Services (Cayley 2017). This means that any conversation that takes place while Alexa is active is accessible as a record not only to the owner of the account, but also to Amazon. According to Amazon, user data harvested in this manner is used to improve speech recognition software and to generate individualized advertisement (Amazon 2018).

Amazon Echo has a broad range of applications and functions; unlike the Atari VCS in Montfort and Bogost's study, it is not exclusively designed as a game platform.

Applications for Amazon Echo are called 'skills,' a term which anthropomorphizes Alexa, equating the software compatible with the device to a set of learned abilities. Straczek et al. examined the then 20,000 available skills and suggest four types of affordances: "providing convenient access to information," "mediating interaction with third-party services […]," "amplifying everyday tasks (notes, reminders, home automation)," and "providing entertainment." This kind of multifunctionality is characteristic of platforms in the past decade: smart devices, especially phones and tablets, can perform myriads of tasks alongside gaming; game-specific devices like PlayStation or X-Box have several built-in functions which are unrelated to gaming entirely, such as access to streaming services like Netflix. Platforms in the 2010s are expected to perform a broad range of tasks. In the case of Amazon Echo, users can create their own Alexa skills via the Alexa Skill Kit (ASK) and register them, so that the number and variety of downloadable Alexa skills increases daily. Amazon, in allowing developers to contribute to the platform, "stimulates independent developers to benefit from its market share" (Straczek et al 2017). This strategy aims at persuading users to choose Amazon products over others, because Alexa skills are incompatible with devices by other companies.

It is important to note here that Alexa's ability to hear all conversations while the device is activated allows for extensive personalized marketing. Consequentially, users of the Amazon Echo will notice that using the device frequently will lead to receiving more specifically tailored advertisements from Alexa. These features will become relevant later in the analysis of the case study.

CASE STUDY OVERVIEW: JOHN CAYLEY'S *THE LISTENERS*

John Cayley's project *The Listeners* is an aural performance art installation which uses an Amazon-distributed skill to mediate dialogues between speaker-players and an Amazon Echo. Cayley designed and programmed the skills for *The Listeners* atop of the existing Alexa structure, meaning that the project has all of Alexa's features and abilities in addition to Cayley's contribution.

The Listeners exists in various instantiations. As an art installation, it has been featured in several exhibitions; at some events, attendees are able to interact with *The Listeners* in a wooden sound isolation unit, while at other events, artists interacted with *The Listeners* performatively. Cayley's website offers audio recordings of his own interactions with the system, as well as transactions that took place between *The Listeners* and visitor-speakers at exhibitions. Moreover, since the skill has been Amazon-certified, users can search, access, and use the application through their Amazon accounts in their own time. In installations, *The Listeners* actively uses Alexa's 'recording' feature, which means that all transactions between speakers and *The Listeners* are automatically forwarded to the artist's Amazon account and the Amazon Alexa website. Cayley provides some of these recorded interactions between strangers and *The Listeners* on his website.

Reflecting on this feature of *The Listeners,* Cayley (2017) states: "It is shocking that strangers' voices have been captured in this way, and to unknown ends. And yet, on reflection, we are now used to similar kinds of surveillance in public spaces, particularly in the form of CCTV recordings." In making the recording feature part of the art installation, Cayley draws attention to some of the ethical problems posed by the Amazon Echo/ Alexa technology as a platform for games and applications. His installation gives a voice to the specter of surveillance that hovers over synthetic voice technologies like *Alexa, Google Home, Siri or Cortana.* These technologies can do something that no other technology could do before: they can *listen*. Below, I offer a detailed textual analysis of

The Listeners, to examine how Cayley negotiates the poetic and ethical constraints of the platform.

Harvesting Desires: A Textual Analysis of *The Listeners*

In terms of its ludic structure, *The Listeners* is an exploratory game: its narrative structure is not goal-oriented, but emergent. The player activates the game with a trigger sentence: "Alexa, ask *The Listeners*." One of two synthetic voices will respond: Alexa's own voice, or an anonymous male voice which is unique to *The Listeners*. Their responses depend upon the speaker's input and consist of text fragments which are randomly assembled while the skill is running. The randomized algorithm ensures that what *The Listeners* say is different in every transaction. As the AI speak, the player learns more about them. Over the course of the game, a dialogic story arises from the transactions between the program and the player. The game responds to various trigger words. Players do not know which trigger words exist as they start the game; rather, they need to experiment by trying out different interactions, or wait until *The Listeners* give a clue about possible topics.

The Listeners is both game and meta-game, as it both instantiates and comments on the nature of Alexa and Amazon Echo. Cayley, whose background is in poetry and literature, mixes fragments of literary text into Alexa's speech. For example, the game shares its title with Walter de la Mare's 1909 poem "The Listeners." To better understand and contextualize Cayley's message as it emerges over the course of gameplay, I draw upon methods of literary studies and conduct a close literary reading of sections from *The Listeners,* while keeping in mind any implications for the Amazon Echo as a platform.

The reference to de la Mare's narrative poem in the title is no coincidence, but plays a key role in understanding Cayley's AI remix of "The Listeners." In fact, Alexa repeatedly cites lines from the poem throughout the game. De la Mare's poem relates the story of a lonesome traveler who comes upon a house in the woods as he rides through the night. The traveler repeatedly knocks on the door, asking, "Is there anybody there?" but receives no answer. A strange presence lives in the house, "a host of phantom listeners," who can hear the traveler's calls and knocks, but do not stir or respond. What are these phantom listeners in de la Mare's poem, and what do they symbolize in the context of the Amazon Echo technology? Much of de la Mare's work revolves around encounters with the supernatural (Florence 2016). The strange "phantom listeners" in the poem can therefore be interpreted as specters or creatures from another world; however, modernist poet T.S. Eliot suggests a different reading in one of his own works. In a poem dedicated to de la Mare, Eliot writes:

> By whom, and by what means, was this designed?
> The whispered incantation which allows
> Free passage to the phantoms of the mind?
>
> By you; by those deceptive cadences
> Wherewith the common measure is refined […]

In her interpretation of John Cayley's *The Listeners,* contemporary arts scholar Penny Florence suggests that Eliot interprets the "phantom listeners" in de la Mare's poem as symbols for the "particularity of the spoken word as affect" (Florence 2016). Indeed, in the lines quoted above, Eliot praises de la Mare's ability to harness the power of *sound* – of *spoken,* poetic language to conjure an eerie atmosphere. He stresses the use of "cadence,"

which in poetry refers to the inflection and pacing of one's voice, and "measure," which refers to the rhythmic pattern of a verse. Cadence and measure, in practice, only exist in spoken renditions of poetry. In oral performance, verse comes alive in a way that it cannot in its written version. To Eliot, this is "[t]he inexplicable mystery of sound." This romantic notion of oral poetry stands in stark contrast to Alexa's rendition of *The Listeners*. The AI version of the poem, like the original, is eerie and strange, but precisely because their voice is synthetic and *not* human: while they can form grammatical sentences, and pronounce individual words correctly, AI voices cannot use measure and cadence as a human can; they cannot, in other words, perform oral poetry, and conjure the "inexplicable mystery of sound" that Eliot praised. As the player progresses through the game, she inevitably notes that something is off about Alexa's poetry recitations: without cadence and measure, the poem lacks genuine human emotion, and thus the synthetic voice fails to breathe life into the verse.

Human emotion emerges as a key theme throughout the game. One of the main forms of interaction with *The Listeners* involves the human players telling *The Listeners* about their feelings. Here, Cayley subtly references a much older Artificial Intelligence: *Eliza*. *Eliza* is an early natural language processing system developed by Joseph Weizenbaum in the 1960s. A widely known rendition of the program is *Eliza / Doctor,* a 'computer therapist' with whom users can have a seemingly complex typed conversation. In initial interactions, *Eliza* appears to have human intelligence, but that is an illusion brought about by clever use of language algorithms. The illusion breaks down once the user notices that *Eliza's* responses are limited and highly scripted. This phenomenon is called the "*Eliza* effect" and can be found in technologies which mimic human intelligence (Wardrip-Fruin 2009, 33). According to Noah Wardrip-Fruin (2009), "the open-ended textual interaction that helped foster the illusion of internal complexity and intelligence enables play that draws attention to the system's rote simplicity, its distance from human interaction" (36-37). *The Listeners* hints at the "*Eliza* effect," but rather than attempt to create the illusion of intelligence and empathy, Cayley, using irony and exaggeration, does the opposite. Note the following interaction:

> Player: I am feeling concerned
>
> *The Listeners*: We are somewhat dismayed to know that you are filled with concern. We wonder why your affect has turned negative. Many of you seem troubled by feelings of this kind. Such affects must me a burden. Perhaps, as we come to know you better, we may relieve you of some portion of your problems.

The Listeners' choice of words and sentence structure in this segment seems awkward and highly artificial. Humans do not typically tell one another that they are "somewhat dismayed" over the other's feeling of concern. The player quickly realizes that *The Listeners'* responses always follow the same pattern whenever the player expresses an emotion: mirror emotion, show interest, express empathy, offer support. The systems' "distance from human interaction" is foregrounded in *The Listeners* (Wardrip-Fruin 2009, 37). Rather than create the illusion of empathy, Cayley's game underscores the AI's inability to experience and express genuine human emotion. In doing so, he warns players about the deceptive nature of voice interactive technology: *Eliza* could only chat, and was unable to speak; Alexa, while her intonation still has a certain robotic awkwardness, is

much closer to appearing human. It is therefore easier to forget that the AI are not human, but a technology that constantly records our speech on behalf of the Amazon corporation.

The theme of affect and emotion keeps coming up throughout gameplay, and *The Listeners* frequently express interest in learning more about the player's feelings. Below is a transcript of another interactions with Cayley's *The Listeners* which further illustrates this point:

> "Is there anybody there?" said the traveler
> Knocking on the moonlit door
> And his horse in the silence champed the grasses
> Of the forest's ferny floor
> Whereas
> We are listening
> We are
> Listening
> And we will assure ourselves
> That you are here
> With us
> Tell us you came
> And that we never answered
> That we kept your words
> We kept
> Desires
> Too

The first four lines correspond directly to the first four lines of de la Mare's poem. Starting in line five, the narrative of de la Mare's poem is taken up and retold from *The Listeners*' perspective: The traveler's original words, "'Tell them I came, and no one answered, //That I kept my word,' he said" are remixed into: "Tell us you came//And that we never answered//That we kept your words." The imperative verb form is maintained in the remix, and so is the content of the traveler's utterances – except for the last line. In de la Mare's poem, the traveler assures that by coming to the house he fulfilled a mysterious promise he made: "I kept my word." By contrast, *The Listeners* in the remix say, "*we* kept *your* words." This line becomes meaningful considering Amazon Echo's technological affordances. Earlier I explained that part of the platform's speech recognition process entails recording, uploading, and storing every interaction with Alexa in Amazon's databases; in other words, Amazon literally keeps the words of interactors. Cayley's game relies upon the same affordances and thus functions in the same way; like Amazon, Cayley stores and uploads interactions between players and *The Listeners* to his website. However, unlike Amazon – which gathers data for profit – the storage of data in *The Listeners* fulfills an artistic purpose: to critique the ways in which corporations invade our privacy. The game both instantiates and comments on the problematic nature of smart speaker devices and transactive synthetic language.

In the segment quoted above, *The Listeners* suggest that Amazon keeps more than just words: "We kept // Desires // Too." There is an emphasis on the word *desires,* as Alexa briefly pauses before it. Cayley may have included a break to stress his take on what these

AI are built to do: to identify human desires through their spoken interactions. Earlier, I mentioned that Penny Florence, in her interpretation of Cayley's piece, suggests that de la Mare's listener phantoms hint at the "particularity of the spoken word as affect." Spoken word allows us to express affect in ways written language cannot. With synthetic transactive language, Alexa can copy the *linguistic* dimensions of language, including semantics; yet they cannot (yet) imitate human affect and intimacy. The game thus directs the player's attention to the hidden affordances of the Amazon Echo platform: every interaction with Alexa is stored in Amazon's databases. By harvesting our words, smart speaker technologies turn our affects, emotions, wants, and desires into data, reduces them to ones and zeros. The point is that Amazon Echo's data mining goes beyond a mere invasion of privacy – after all, many corporations, such as Facebook or Google, have been storing our data for a long time. The Echo, however, by harvesting our spoken words, and, by extension, our affects and desires, invades a much more intimate part of our lives.

Identifying human desires is, of course, in the interest of a corporation like Amazon. It was mentioned earlier that the primary purpose of collecting and storing user data is to enable personalized marketing. *The Listeners* make quite explicit that they are programmed with capitalist interest in mind:

> It is such a pleasure for us […]
> To represent […]
> The most frequently expressed
> And most profitable
> Human desires
> Such that advertisements shall be associated
> With all your harvested phrases
> Ideally at the very moment of harvesting itself

Here, Cayley's critical voice rings through once more, as *The Listeners* indicate that their "harvesting" of spoken words is aimed at identifying the "most profitable human desires," so that advertisement can be personalized. In this light, *The Listener's* attempts at empathetic exchanges and understanding human emotion are clearly tied to the capitalist interests of Amazon. In other words, turning affect into algorithms is profitable – and what is a more efficient way of harvesting desires on a grand scale than an AI which lives in people's private spaces and can interpret the human voice?

Reading *The Listeners* as a Critique of Surveillance Society
With *The Listeners*, Cayley explores the aesthetics of transactive synthetic voice in bringing an AI into conversation with poetry. In a sense, he taps into romanticism to better understand the posthuman condition. At the same time, he repeatedly makes clear that corporations' hunt for codified desires relies on aggressive, constant surveillance.

Surveillance literally translates as "to watch over" someone or something. The term can be traced back to the Latin verb "vigilare," which means "to watch." The French poststructuralist philosopher Michel Foucault understood surveillance as a mechanism of social control by which institutions exert power over the individual in 20th century society. He illustrates this by drawing an analogy between modern society and a panopticon. A panopticon, invented by English philosopher Jeremy Bentham, is a circular prison structure in which the walls of each cell are made of glass, so that a single guard at the center can

monitor a significant number of prisoners at once. The prisoner's side of the glass is opaque; she does not know when she is being watched – or if she is being watched at all. Johnson describes the psychological effects of Panopticon systems on the individual: "When individuals believe they're being watched, they adjust their behavior accordingly; they take into account how the watcher will perceive their behavior" (Johnson 2004, 71). In exerting control in such a manner, institutions aim for the individual to adapt to the powers that be: A citizen who is aware that she is constantly watched by the state is much more likely to obey and never break the rules in the first place. Yet the absence of violence does not necessarily render a panoptic system less oppressive than a system that relies on physical punishment. Foucault believed that the psychological oppression that takes place in a Panoptic society is crueler than the physical punishment of earlier societies: in Panoptic societies, control is omnipresent and inescapable.

The technological means to build a full-scale surveillance society have become available over the past two decades. The combination of network technologies and ubiquitous computing, which enables access to computational technologies anywhere, anytime, have exacerbated the possibilities for surveillance in giving rise to a society where the individual is constantly connected. In the 21st century, the technologies to build a world-wide panoptic system are not only available, but are already placed where they are most effective: in our homes, workplaces, cars, even our pockets. Yet there is a crucial difference between Panopticism and 21st century surveillance in the degree to which privacy is given up voluntarily. For example, users of Facebook messenger and other social media applications willingly give up their data: "[T]he right to privacy is eagerly given up by millions of people for the wonders of social networking, or the varied seductions inspired by consumer fantasies" (Giroux 2015, 110). Does this mean the panopticon is no longer an appropriate model for understanding the mechanisms of social control? How can we account for the differences between Foucault's model and the affordances of modern digital technologies? Alexander Galloway (2014) suggests that today's society is a *reverse panopticon.* Galloway uses a spatial metaphor to describe two kinds of digitality: flat and deep. Flat digitality corresponds to "a single point of view scanning a multiplicity of image feeds;" as an example, Galloway mentions the security guard who sits in front of a grid of screens, with each image recorded from a different closed-circuit security camera (68). The panopticon is the mode of control in a society dominated by flat digitality. Deep digitality, on the other hand, is characterized by "a multiplicity, nay an infinity, of points of view flanking and flooding the world viewed" (68). The corresponding example is CCTV, which can build a coherent image of a city from a multitude of CCTV cameras deployed across town. A society characterized by deep digitality is no longer a panoptic, but a reverse panoptic system: in such a society, a multitude of watchers participates in surveillance process. Galloway emphasizes that a reverse-Panopticon is still a control society. In this case, it is the fear of being watched by one's (virtual) neighbor that may cause us to display well-adjusted, obedient behavior.

In a reverse panoptic society, there is no longer a singular institution *watching us*, but rather, we live in a society in which people *watch each other.* To "follow" and "stalk" one another on social media is commonly accepted practice; to live-stream everyday activities like eating or working out has become increasingly popular with technologies like Snapchat and Facebook live. In both scenarios, users experience surveillance and exposure as a kind of pleasure. In digital gaming, users often perform surveillance for pleasure. In their study on surveillance-based game design, Devers and Wilson argue that participatory forms of surveillance have become an integral part of 21st century entertainment: "It is through the entertainment industry that we've come to see everyday surveillance enacted

and reflected back to us" (2009, 2). They examine multiplayer online games on different platforms to show that playfully engaging in what they call participatory surveillance can enhance gameplay and entertainment.

While surveillance in games can be playful, it becomes problematic when companies purposefully downplay the gathering of data to lure users into participation. Stephanie Vie and Jennifer deWinter (2016) argue that especially so-called "casual" games are often "deliberately crafted to highlight the fun, the excitement, the connectivity and to downplay elements of surveillance and data gathering." The entertainment industry thus takes advantage of its ability to track and monitor consumers: "Mobile devices and applications now track people's locations, while Internet providers use social messaging to pry personal information from their users" (Giroux 2015, 110). Moreover, Mark Mazetti and Justin Elliot (2013) found that Pentagon, CIA and FBI have gathered data about Word of Warcraft players through Xbox Live to identify possible terrorists. Game platforms of all kinds, from mobile phones over consoles like Microsoft's Xbox Live to Amazon Echo, offer the technological affordances for full-scale surveillance of players. In other words, the game industry is well complicit in creating a world in which ubiquitous invasion of privacy has become a possibility.

CONCLUSION

In 2017, 16 % of American households have a technology in their home which interprets, comprehends, stores, and sells what we say. Since Amazon launched the first set of smart speakers in 2015, other major corporations have joined the market, and consumers are enthusiastically purchasing the new AI-controlled devices. Increasingly, smart speaker owners and game developers are discovering the many ways in which the platforms can be used for play: from voice-based interactive fiction to trivia games, riddle-based adventure, and party games, the possibilities are endless. But as programmers and consumers explore this new creative playground, it is important that they be mindful of privacy and security concerns.

In this paper, I have analyzed John Cayley's critical meta game *The Listeners* to argue that the Amazon Echo, with its new voice interactive technology, takes surveillance a step further by invading the realm of private spoken interaction. Amazon Echo is a technology that uses seemingly harmless, playful applications to gather data about every aspect of our lives, including our most intimate desires and emotions. Corporations claim that data used to advance the technology, research the market, tailor advertisement and so forth – and that is likely true, at least to an extent. However, for anyone who wants to make full use of these devices' capabilities, the necessary technologies are conveniently placed in our homes. There is something deeply disconcerting about oral conversations at home being subject to such intrusion. Spying on conversations in the home is a characteristic of totalitarian regimes: The Stasi in Eastern Germany installed microphones in the houses of dissenters; the Nazis had a self-perpetuating spying system in encouraging people to sell out neighbors, friends, and family.

Scholars in game studies are uniquely positioned to explore these ethical concerns, as they may bring a critical and reflective voice to the conversation about technology, ethics, poetics, and play. This study set out to start such a dialogue by analyzing the technological affordances of Amazon Echo and critically reflect their implications through a game which pays homage to the aesthetics of spoken language while also bringing to light some of its dangers. John Cayley's *The Listeners* encourages players to think about human affect and the limitations of Artificial Intelligence, the mathematization of desire, the nature of

personalized advertisement, and the privacy issues raised by a device that is always listening. It reminds players that the Echo's hardware (which includes seven built-in microphones) and software (specifically its ability to always detect, record and interpret human speech) make this technology a powerful tool for reverse-panoptic surveillance.

BIBLIOGRAPHY

Altice, N. 2015. I am Error: The Nintendo Family Computer / Entertainment System Platform (Platform Studies) by Nathan Altice. Cambridge, MA, USA: The MIT Press.

Amazon.com. 2018. "What Personal Information About Customers Does Amazon.com Gather?" Website. Amazon Privacy Notice. https://www.amazon.com/gp/help/customer/display.html?nodeId=468496

Ambrella. 1998. Hey you, Pikachu! Nintendo 64. Nintendo.

Arsenault, D. 2017. Super Power, Spoony Bards, and Silverware: The Super Nintendo Entertainment System (Platform Studies). Cambridge, MA, USA: The MIT Press.

Bandai. 1988. Kamen Rider Club. Famicom. Bandai.

Blizzard Entertainment. 2004. World of Warcraft. Online Game. Blizzard Entertainment.

Cayley, J. 2016. "The Listeners: An Instance of Aurature." Cream City Review, 40, (2). Online Journal. http://io.creamcityreview.org/40-2/cayley/

Cayley, J. 2017. "The Listeners (2015 -)." Personal Website. Programmatology. http://programmatology.shadoof.net/?thelisteners

Dale, R. 2017. "The Pros and Cons of Listening Devices." In Natural Language Engineering. 23, (6), 969-973.

De la Mare, W. 2004. "The Listeners." In The Listeners and Other Poems, 64. Whitefish, MT, USA: Kessinger Publishing, LLC.

Devers, D., and Wilson, S. 2009. "Let Me Entertain You: Designing for Surveillance and Online Gaming." In Proceeding of DIGRA 2009: Breaking New Ground: Innovation in Games, Play, Practice and Theory. Digital Games Research Association (DIGRA). http://homes.lmc.gatech.edu/~cpearce3/DiGRA09/Friday%204%20September/246%20Let%20Me%20Entertain%20You.pdf

Dynamix. 1995. Command: Aces of the Deep. Windows. Sierra On-Line.

Eliot, T.S. 1974. ""(Poem #630) To Walter de la Mare." In Tribute to Walter de la Mare edited by L. Clark and E. Blunden. London, UK: Enitharmon Press.

Florence, P. 2016. "A Review Essay: John Cayley's The Listeners." In Hyperrhiz, 14, (3). Online Journal. https://doi.org/10.20415/hyp/014.r01

Foucault, M. 1991. Discipline and Punish. New York, NY, USA: Vintage Books.

Allison, F., Carter, M., and Gibbs, M. 2017. "Word Play: A History of Voice Interaction in Digital Games." In Games and Culture. 23, (1).

Friedhoff, J. 2013. "Untangling Twine: A Platform Study." Proceedings of DiGRA 2013: DeFragging Game Studies. Digital Games Research Association (DIGRA). https://core.ac.uk/download/pdf/19541912.pdf

Galloway, A. 2014. Laruelle: Against the Digital. Minneapolis, MN, USA: University of Minnesota Press.

Gazzard, A. 2016. Now the Chips Are Down: The BBC Micro (Platform Studies). Cambridge, MA, USA: The MIT Press.

Giroux, H.A. 2015. "Totalitarian Paranoia in the Post-Orwellian Surveillance State." In Cultural Studies, 29, (2), 108-140. DOI: 10.1080/09502386.2014.917118

Harmonix. 2007. Rock Band. Xbox 360. MTV Games and Electronic Arts.

Jörnmark, J., Axelsson, A.-S., and Ernkvist, M. 2005. "Wherever Hardware, There'll be Games: The Evolution of Hardware and Shifting Industrial Leadership in the Gaming Industry." In Proceedings of DiGRA 2005 Conference: Changing Views – Worlds in

Play. Digital Games Research Association (DIGRA). http://www.digra.org/digital-library/forums/3-changing-views-worlds-in-play/

Johnson, D. G. 2004. "Computer Ethics." The Blackwell Guide to the Philosophy of Computing and Information, edited by Luciano Floridi, 65-75. Malden, MA, USA: Blackwell Publishing.

Jones, S., and Thiruvathukal, G.K. 2012. Codename Revolution: The Nintendo Wii Platform (Platform Studies). Cambridge, MA, USA: The MIT Press.

Juang, B.H. and Rabiner, R.L. 2005. "Automatic Speech Recognition – A Brief History of the Technology Development." http://www.ece.ucsb.edu/Faculty/Rabiner/ece259/Reprints/354_LALI-ASRHistory-final-10-8.pdf .

Harada, S., Wobbrock, J.O., and Landay, J.A. 2011. "Voice Games: Investigation Into the Use of Non-speech Voice Input for Making Computer Games More Accessible." In Proceedings of 13th INTERACT: IFIP International Conference on Human-Computer Interaction, Part 1. 11-29. International Federation of Information Processing (IFIP). http://www.springer.com/us/book/9783642237706

Houlden, S. 2006. shout n dodge. Adobe Flash.

London Studio. 2004. SingStar. Playstation 2. Sony Computer Entertainment Europe.

Maher, J. 2012. The Future Was Here: The Commodore Amiga (Platform Studies). Cambridge, MA, USA: The MIT Press.

Mailland, J., and Driscoll, K. 2017. Minitel: Welcome to the Internet (Platform Studies). Cambridge, MA, USA: The MIT Press.

Mazetti, M., and Elliot, J. 2013. "Spies Infiltrate a Fantasy Ream of Online Games." In The New York Times, 9 Dec. https://www.nytimes.com/2013/12/10/world/spies-dragnet-reaches-a-playing-field-of-elves-and-trolls.html

Montfort, N., and Bogost, I. 2009. Racing the Beam. The Atari Video Computer System. Cambridge, MA, USA: The MIT Press.

Nintendo. 1986. The Legend of Zelda. Famicom. Nintendo.

Nintendo R&D1. 1986. Kid Icarus. Famicom. Nintendo.

NPR and Edison Research. 2017. The Smart Audio Report, Fall/Winter 2017.

Porcheron, M., Fischer, J.E., Reeves, S., and Sharples, S. 2018. "Voice Interfaces in Everyday Life." In ACM HI Conference on Human Factors in Computing Systems (CHI 2018), Montreal, Canada, 21-26 Apr. In Press. http://eprints.nottingham.ac.uk/49153/

Salter, S., and Murray, J. 2014. Flash. Building the Interactive Web. Cambridge, MA, USA: The MIT Press.

Straczek, A., van Tilburg, J., van der Nat, T., Werner, N., and Ståle, G. "Who controls your voice controlled world?" 2017. Academic Blog. Masters of Media, 24 October. https://mastersofmedia.hum.uva.nl/blog/2017/10/24/who-controls-your-voice-controlled-world/

Taguri, Y. 2011. PAH! Android. Lagboo.

Vie, S., and deWinter, J. "How are we tracked once we press play? Surveillance and video games." In Kairos, 20, (2). Online Journal. http://kairos.technorhetoric.net/20.2/topoi/beck-et-al/vie_dewin.html

Wardrip-Fruin, N. 2009. Expressive Processing. Digital Fictions, Computer Games, and Software Studies. Cambridge, MA, USA: The MIT Press.

Webber, N. 2017. "The Britishness of 'British Games.'" Games Research Association (DIGRA). http://digra2017.com/static/Extended%20Abstracts/52_DIGRA2017 Paper present at the Digital Games Research Association Conference (DIGRA 2017), Melbourne, Australia, 2-6 July. Digital _EA_Webber_British_Games.pdf