Article

new media & society
1–18
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1461444820923679
journals.sagepub.com/home/nms
SAGE

Preparing for smart voice assistants: Cultural histories and media innovations

Justine Humphry and Chris Chesher
The University of Sydney, Australia

Abstract

Smart voice assistants have become popular thanks largely to their default naturalistic female voices and helpful personae. In this article, we trace changes in robot voices in popular culture and explain how this history influenced the voice design of smart voice assistants. Our research draws on cultural analysis of Hollywood and international films, television and literature, and observations from our personal experiences with voice assistants. We argue that designers of devices like the Google Home and Amazon Echo inherited a cultural imaginary of alien and dangerous robots with artificial voices and personalities. Manufacturers leveraged techniques of modality, personae and invocation and pre-existing social connotations of the voice to create positive associations of these devices in the home. We conclude by arguing that smart voice assistants are new media innovations prepared for consumers through pre-domestication and represent an emerging regime of power and influence based on technologised voice interaction.

Keywords

Artificial intelligence, domestication, gender, invocation, media innovation, modality, personae, robots, smart voice assistant

Introduction

When the Amazon Echo was launched in November 2014 it was not entirely strange. Apple had already released Siri, and Microsoft had Cortana. We were used to satellite navigation systems that talked, and 'interactive voice response' telephone agents with which we had stilted conversations. However, the familiarity with artificial voices ran

Corresponding author:
Justine Humphry, Department of Media and Communications, The University of Sydney, Camperdown, NSW 2006, Australia.
Email: justine.humphry@sydney.edu.au

much deeper. The invention of the phonograph in the 19th century was a milestone in the cultural familiarisation with technologies that speak and make sound, disrupting the assumed connection between the voice of the human speaker and the presence of his or her body (Katz, 2010). Through everyday experiences with audio recordings, telephony and radio, people adapted to the psychological effect known as schizophonia, a psychic disturbance resulting from the separation of sound from its original source (Schafer, 1969). Technologies for speech synthesis, and then talking robots, made regular appearances in science fairs, theatre, literature, cinema, television and computer culture (Pieraccini, 2012). Conventions for how robots' voices should sound and talk, and what they should speak about, have developed since early in the 20th century in popular culture and scientific experimentation (Cox, 2018).

We argue that these histories have informed the emerging discipline of voice user interface (VUI) design applied in the current generation of smart voice assistants on phones and smart speakers. At the same time, we analyse the media innovations and strategies of United States–based technology firms that prepare these devices to support their goals of global expansion. The use of naturalistic middle-class, female voices and personae positions speakers as friendly participants in everyday family routines. As argued by Woods (2018), this allays fears of surveillance capitalism (Zuboff, 2018) associated with the projects of companies such as Amazon and Google. As we will show, this also works within and against a cultural history of robot voices that play on the social and gendered connotations of voice (Phan, 2017).

We begin by examining the emergence of the VUI as a paradigm of human–computer interaction and survey the literature on voice assistants and theories of voice and sound. We examine a number of archetypes of talking robots and computers from Hollywood and international screen culture and explore how these manifestations of the cultural imaginary have informed the design and cultural reception of smart voice assistants. We introduce the concepts of modality, personae and invocation to help explain the selections and exclusions in the voice design of smart voice assistants, and how these shape user interactions in the dominant US smart speaker brands: Amazon, Google and Apple. We apply Van Leeuwen's (1999) concept of modality to show that the voices of smart voice assistants are designed to seem more naturalistically 'human' to avoid the historically negative associations with robotic voices and artificial intelligence. The personae of voice assistants have equally been engineered to adopt norms of gender, race and class to reduce anxieties about their potential to exceed their roles as loyal helpers and cross the boundary into the monstrous. We also analyse the distinctive acts of invocation that characterise these devices as a media form (Chesher, 2001, 2004).

Finally, we analyse how voice design fits into corporate strategies to prepare the smart voice assistant for their encounters with consumers in the home. We will argue this follows an expansionist imperative that drives the design and marketing of this new category of media devices (Goulden, 2019). We have described this process of managing the meanings of the voice and design of these digital assistants as pre-domestication (Saariketo, 2018; Silverstone and Haddon, 1996, 1998). Their strategy involves refining smart speaker voices to a set of cultural stereotypes while making invisible the social, technical and political processes by which these design decisions are made. This approach anticipates anxiety and resistance from consumers to the surveillant potential of these technologies, while keying into more positive associations of robots and the automated home of the future. Methodologically, we adopted a cultural history approach (Davis, 2009), documenting and analysing synthetic voices in cinematic and televisual representations, advertising, technical systems and associated documents. Following Uotinen's (2010) lead on autoethnography as a method for studying cultural aspects of technology through the researcher's personal life, our research also draws on observations of living with voice assistants. We draw on theories of voice, speech acts, orality and media (Austin, 1962; Chesher, 2004; Guattari, 1992; Nass and Brave, 2005; Ong, 1982[1982]; Searle, 1969; Van Leeuwen, 1999) to conceptualise and interrogate the technocultural construction of voice assistants and VUIs as a new model of human–computer interaction.

A new model of human–computer interaction

The VUI is a model of human–computer interaction that enables users to interact with technologies using their voice to invoke spoken commands (Mortensen, 2019). VUIs can be found in a wide range of devices, including mobile phones, satnavs and televisions, and while they build on a long history of scientific developments in speech recognition, their recent growth and the industry excitement around these has come about through the global uptake of smart speakers with voice-controlled virtual assistants such as Apple's Siri, Google Assistant and Amazon's Alexa. By the end of 2018, more than 100 million households worldwide had acquired smart speakers (Kinsella, 2018). A 2019 consumer adoption report found that Australia is an especially rapid adopter of smart speaker technology with over 10 million devices sold in the country by March 2019 so that about 5.7 million or 29.3% of the population own a device (Kinsella, 2018). One industry projection is that by 2021, there will be more voice-activated assistants on the planet than people (Statista Research Department, 2020). As a United Nations (UN) report claims, these trends 'signal that the way people interface with technology is in the midst of a paradigm shift from text input and output to voice input and output' (West et al., 2019: 93).

A driver in the rise of VUIs is the powerful belief that speech is the most natural form of human communication (Phan, 2017). Advocates of VUIs idealise speech as the most advanced, natural and transparent means for communicating with computers. One developer's video, introducing the programmable features of Amazon's Alexa, adapted the well-known diagram of the stages of human evolution to position voice interaction as the pinnacle of computing, while representing desktop computers and tablets as more primitive technologies (Cutsinger and Akersh, 2018). Mortensen (2019) expounds on the superiority of voice for design at *The Interaction Design Foundation*, citing Nass and Brave, who wrote in their book *Wired for Speech* that 'Speech is the fundamental means of human communication. Even when other forms of communication – such as writing, facial expressions, or sign language – would be equally expressive, (hearing people) in all cultures persuade, inform and build relationships primarily through speech' (Nass and Brave, 2005: 1).

Voice assistants recall some of the cultural and psychodynamic features of oral culture identified by Walter Ong (2002[1982]). Ong argues that participants in pre-literate primary oral cultures communicate in the present, operate through sound (rather than vision), and perform in the immediate interactive presence of participants (Ong, 2002[1982]). He argues that some features of oral culture have reappeared in a world dominated by writing and print. He referred to this as 'secondary orality' pointing to the telephone, radio and television as examples of media that have features of oral culture such as the news anchor who performs as a story teller. VUIs are distinctively different from technologies of writing such as the keyboard and mouse, manifesting a version of secondary orality. Where web searches are forms of writing, voice assistants are oral. Where web searches deliver multiple textual results, a voice assistant typically provides a brief spoken response. Where different web sites represent a range of different voices, the voice assistant typically maintains a consistent voice and persona with which the user can establish an ongoing relationship.

The social character of voice complicates the apparent naturalness of voice as a medium of human communication. Voice immediately reveals the identity and emotion of the speaker. Voices are gendered. They give away ethnic, regional and subcultural identity as established in cultural studies, sociology and linguistics (Bucholtz and Hall, 2009). They give off emotions such as depression, fear or excitement. Designers of the voices of voice assistants face significant challenges in creating voices that belong within the social spaces in which they speak (Nass and Brave, 2005). The voices of the most popular voice assistants are localised to the region in which they are marketed. In Australia, Siri, Google Assistant and Alexa speak in middle-class Australian accents. Until recently, multiple Google Assistant voices have only been available in the United States, and only in English (Roettgers, 2019). Phan (2017) argues that assistant voices are engineered to engender trust and apparent transparency.

Critical scholarship on smart speakers has focussed on their technocultural construction, and particularly the cultivation of a female voice and persona as the default. Bergen (2016) points out that it is significant that these voices are disembodied and 'devoid of that leaky, emotive quality that we have come to associate with the feminine body' (101). Strengers and Nicholls (2017) see the voice assistant as a 'wife replacement' (p. 6). Woods (2018) argues that invoking stereotypes of caretaker, mother and wife softens any anxieties from consumers about the surveillant implications and consumerist imperatives of these devices and systems. Hui and Leong (2017) argue that the mobilisation of anthropomorphism, sociability, convenience and humour aims to habituate users to these devices. Phan (2019) argues that Alexa is an aestheticisation and decontextualisation of the domestic servant. In a 'Think piece' in a recent UN report (2019) West et al. (2019) argue that companies recognise that their voice personae are their representatives, and as such need to create a favourable impression. In this context, female voices are preferred, but, this 'has less to do with sound, tone, syntax and cadence, than an association with assistance' (p. 98). The preference for female voices, he argues, also relates to their appeal to the predominantly male (and presumably heterosexual) population in the tech industry developing these devices.

In this article, we add to this literature with our argument that the selection of the voice for voice assistants is informed by a cultural history of artificial voices that serves as a resource for managing the meanings and consumption of designed voices. In the history of screen robots and talking machines, the voices and personae are often exaggerated stereotypes: menacing antagonists, seductive love objects, comic naïfs, father and mother figures, innocent children, loyal pets, children's companions and obsequious servants. Each of these robot archetypes is characterised by a voice that diverges from the naturalistic, and carries meanings that constitute the robot or computer as 'other'. These caricatured voices contrast with voice assistants, which designers aspired to make as 'natural' as possible, at first in their default middle-class female voice. Siri, Alexa, Cortana and the Google Assistant were developed at a time when the voices of robot characters in popular culture were becoming increasingly naturalistic and these devices were developed in relation to these trends.

Histories of the machinic voice

Engineers have long experimented with speaking machines, with varying degrees of success. Kempelen's 18th-century machine modelled on the lungs and the vocal tract could make only a limited range of speech-like sounds (Linngard, 1985). In the 1920s, Bell Labs engineers built the Voice Operation Demonstrator, or Voder: a complex device with keyboards, wrist bar and foot pedal that electronically generated around 20 sounds that a talented operator could 'play' to create comprehensible speech (Eschner, 2017). It took more than a year of training, and hours of daily practice to become a competent performer with the device. Of 300 female telephone operators on their staff, Bell chose 24 with the most aptitude to train up to use the Voder in trade shows and demonstrations (Smith, 2008). Even with this skill, artificial speech was slow and deliberate, composed of various manipulations of generated waveforms. The Voder's nickname was 'Pedro', making him not only male, but foreign (Smith, 2008). A newsreel film on YouTube (Roemmele, 2016) demonstrating the Voder at the 1939 World Fair in New York reveals a telling hierarchy of voices. Pedro's voice is clearly male, while the operator's hands are female and White, and she is instructed by a male engineer. The woman's partly disembodied voice relegates her to being an extension of the machine despite her virtuosic expertise required for its operation. Meanwhile, the unmarked (and unremarked) Whiteness of the narrator reinforces the masculine voice of authority.

The Voder demonstrations, which received plenty of media coverage, helped establish the norms for robotic speech that echoed the pre-existing social authority credited to the male voice (Lanser, 1992). This convention would be taken up and renegotiated in popular culture in the decades that followed. It is notable that the term 'robot' was first coined and popularised in the 1920s play RUR, in which actors perform with robotic voices (Čapek et al., 1928). Robotic voices were 'robotic', but rarely entirely monotone. Some of the earliest electronic speech synthesisers used prosody: variations in rhythm and pitch, placing stress on certain syllables to achieve different meanings and suggest personality (Cox, 2018). The staccato robotic style of speaking was apparent in the voice performance of Westinghouse's Elektro Moto Man robot, a hybrid of vaudeville theatre and engineering that used eight turntables to play back recorded speech to crowds at the New York World's Fair in 1939 (Rydell, 1990). When digital computers automated speech production in the 1960s, also in Bell Labs, they spoke with a similar distinctively machinic masculine voice, echoing the voice of the engineers who built them.

The archetype of the robotic voice established by the mid 20th century in popular culture was distinguished from human voices through its low modality, which made robots seem machine-like, inhuman and ‘other’. In linguistics, modality relates to the degree of certainty of a proposition. A sentence such as ‘The earth is flat’ has high modality: it expresses certainty. On the other hand, ‘I believe the earth is flat’ introduces some doubt, and lower modality, distancing itself from the truth claim. Theo Van Leeuwen (1999) extended the concept of modality to refer to the various sonic changes that differentiate a naturalistic sound from a processed or distorted one. He argued that processed or distorted sounds have lower modality and therefore less certainty, which helps explain modality in the voice of a robot whose status as a subject is ambiguous. The quality of a robot character’s voice, whether synthesised by a machine or performed by an actor, was made deliberately non-naturalistic in early 20th-century popular culture. Most 20th-century filmmakers followed this convention of distinguishing robot voices from human characters by lowering naturalistic modality in dialogue, performance and sound manipulation, marking out robots and computers as other.

Menacing machines

Where the first publicly exhibited robots in the early 20th century were presented as marvels of futuristic technology, as the century progressed, science fiction robots sounded darker and more sinister. Computers and robots became associated with narratives of science gone out of control (recalling *Frankenstein*), and later with the military industrial complex. In our review of mid-20th-century to early 21st-century Hollywood films, we identified a range of dystopian screen narratives that saw the masculine robotic voice become menacing and potentially violent. Using either speech synthesis with devices like the Vocoder, or electronic processing techniques, filmmakers produced distinctive lower modality sounds that gave the robots the voice of the other (Krapp, 2011). This is heard in many films such as *Forbidden Planet* (1956), the sci-fi horror film *The Collossus of New York* (1958) and the infamous computer HAL 9000 in Stanley Kubrick’s *2001 A Space Odyssey* (1968) where the computer’s allegiance to the mission at the cost of the crew becomes murderous. As the astronaut Dave shuts down the computer, HAL’s voice becomes increasingly distorted and inhuman. A similar theme of the menacing machine is apparent in *Alien* (1979) when crew member Ash is exposed as an android. After a violent struggle in which he is grotesquely damaged, his voice becomes distant and robotic as he recites the directives of the corporation: ‘Bring back life form. Priority One. All other priorities rescinded’. As he breaks down, his voice reverts to the synthesised components that make up machine speech, revealing his ‘true’ artificial nature.

The robotic anti-hero is not always male. Monstrous female robots are typically pathological mother figures whose well-meaning intentions go haywire. In the Disney movie *Smart House*, the smart home, personified as the voice of PAT, turns into a controlling mother who flies into a rage when the family refuses to cede to her maternal demands. In Alex Proyas’ *I, Robot*, the computer VIKI and her robot hoards turn against the people in a maternal effort to protect the future of humanity from itself. In a highly processed robotic voice, she chides, ‘You charge us with your safekeeping, yet despite our best efforts, your countries wage wars, you toxify your Earth and pursue ever more imaginative means of self-destruction. You cannot be trusted with your own survival’. In *Alien*, the mainframe computer onboard the spaceship Nostramo, which does not actually speak, is called ‘MOTHER’. In the 2019 Netflix film *I am mother*, the robot ‘Mother’ speaks with a processed, stern but maternal voice to her ‘daughter’; but it turns out she too is excessively controlling.

Another common characterisation of the female robotic anti-hero, extending the trope of the overprotective mother, is that of the parthenogenetic maternal figure who turns abject (Creed, 1993). In the 2017 Spanish film *Órbita 9*, the orphan Helene is cared for on the Orbiter Space Station by Rebecca, the onboard computer. Rebecca looks after her bio-needs and is her main form of company, with the exception of the occasional visit by a ship maintenance worker. The false warmth and evenness of Rebecca’s vocal ministrations are exposed when Helene discovers she is being kept prisoner on a ship that has never left earth.

As Sofoulis (2001) explains in her analysis of the Star Trek television series, the starship Enterprise functions as a technoworld, an ‘agentic and knowing container with various maternal qualities, especially life support’ (p. 137). Sofoulis points out that the ‘container-mum’ (p. 144) manifests in such infrastructural sounds as background hums, high-pitched electronic beeps and electronic sounds emitted from onboard equipment and displays. The ship-as-space, however, can become a dangerous actor whose intentions are not as they seem (as in the cases of Rebecca in *Órbita 9* and Mother in *I am mother*) or as a monstrous enclosure (in the case of MOTHER in the film *Alien*), in which she is invaded from within, a powerless witness to the deaths of the crew and to Ripley’s efforts to survive. The low modality voices in these films express the figure of the menacing robot, which we will argue informs the selection of vocal modality in the design of smart voice assistants.

Intelligence, ineptitude and humour

Another common trope for cinematic and televisual robots in comic scenes is to present male robot characters as super-intelligent, but socially inept. Robots often speak in an unnecessarily formal and staccato style, and move in a jerky and uncoordinated manner. In the 1956 film *Forbidden Planet*, Robbie the Robot moves awkwardly into the scene, obsequiously bows, welcomes the visitors from the spaceship as ‘Gentlemen’ and boasts that he can speak 187 languages. In the *Star Wars* series, C3PO often brags about his superior facility with languages, but his awkward movements and neurotic personality make him a comic but sympathetic character. Many other screen robots talk with excessive formality, often drawing attention to their superhuman mental faculties. For example, Orac in *Blake’s 7*, KIT in *Knight Rider* and K9 in *Dr Who* use unnaturally technical language with heavily electronically processed voices. In *Hitchhiker’s Guide to the Galaxy*, the robot Marvin the Paranoid Android is a depressive character who complains that although he has ‘a brain the size of a planet’, his capacities are underappreciated. When Professor Stephen Hawking took up a speech synthesiser to communicate, the association of robotic speech with intelligence made the choice of the synthesiser seem entirely logical.

Cinema often depicts robots as naive and unworldly – the fish out of water – such as John Malkovich’s character in *Making Mr Right*. In a shopping centre, the robot played by Malkovich spots an acquaintance ascending an escalator. The acquaintance greets him by observing: ‘It’s a small world’. The robot takes this literally, confirming her statement with the measurements of the earth. In each of these cases, human and robot characters are marked in the actor’s appearance, bodily performance and affected speech. They clearly do not fit in the human world. Kriz et al. (2010) argue that such depictions have led to a general expectation that robots have advanced cognitive abilities but lack social skills.

The *Terminator* franchise plays with vocal modality in the Terminator’s transition from anti-hero in the first film to hero in the second, and from killer robot to honorary human. In *The Terminator* (1984), Arnold Schwarzenegger’s strong Austrian accent, deep voice and staccato delivery serve as appropriately robotic, as typified in the delivery of the famous phrase ‘I’ll be back’. By *Terminator 2: Judgement Day*, when Arnie returns as a sympathetic military android sent to protect the young John Connor, the teenage John spends time trying to teach him how to speak in contemporary US vernacular:

> No, no, no, no. You gotta listen to the way people talk. You don’t say ‘affirmative’, or some shit like that. You say ‘no problemo’. And if someone comes on to you with an attitude you say ‘eat me’. And if you want to shine them on it’s ‘hasta la vista, baby’.

Even non-humanoid robotic characters play with vocal modality. In the film *Interstellar*, the robot character TARS has a synthesised voice inflected with irony, and there is comedic value in the fact that the robot has personality settings, so that users can choose the level of honesty and humour. TARS’ dialogue modality adjusts accordingly, which illustrates how robots might be designed to be emotionally manipulative.

In all these examples, the voice is a crucial vehicle with which robots and other speaking machines express and perform a *persona*. The concept of persona, meaning a fictitious character identity, has a long history, coming from the Greek word ‘prosōpon’ which means mask. The persona functions as the key organising schema in character-driven forms of entertainment including theatre, fictional literature, television and cinema (Sadoski, 1992). VUI developers adopted this concept after recognising its value for getting consumers to identify with their products (Nielsen, 2013). As Dasgupta (2018) argues in a VUI textbook, it helps in ‘making the interaction easier and more natural’ (p. 43). These developments, drawing on a rich theatrical and literary tradition, have underpinned the shift in the representation of the robot towards more complex and sympathetic characters with distinct personalities.

Naturalistic personae

A landmark in this transition to more naturalistic screen robots, or replicants, is the 1982 film *Blade Runner.* The replicants are very hard to distinguish from humans, to the point where only specially trained blade runners are able to identify them. Replicants’ speech is indistinguishable from human characters. If anything, they are more empathetic, eloquent and poetic than humans in that world. In the climactic scene, the replicant Roy Batty stops fighting Deckard, the blade runner pursuing him, and rescues him from falling to his death. Batty becomes a tragic figure in a famous dying monologue in a notably dramaturgical non-robotic voice. Bishop (2014) argues that the naturalistic persona can work as a cinematic device for exploring the boundaries between human and machine, such as the robot that wishes to be human (the ‘Pinocchio Predicament’) or alternately, the robotic villain that rejects humanness.

In the 21st century, screen robots have increasingly tended to become more psychologically complex characters, and this has been accompanied by variations in modality in voice and speech. In the 2015 TV series *Humans*, the standard-issue service ‘synths’ are differentiated modally from a featured group of sentient synths who move and speak more like real humans. Their evolution into sentient beings is marked by the elimination of modal distance, as their gait becomes looser, and their speech takes on features of natural conversation such as pauses, turn-taking signals and other marks of expressiveness. The 2016 TV series *Westworld* also plays with modalities of the robotic voice. In the Westworld theme park, the robots ‘hosts’ frequently transition between different computer-like modes that are manifest as performative modalities. In the world of the theme park, they are usually in character mode: playing a theatrical role in the simulation of the 19th-century US West. When they are backstage, being repaired, the technicians are able to put them into ‘analysis mode’, in which they impassively give feedback on their machine state to trouble-shoot faults. The drama in the first series revolves around the robots’ struggle to reach an awareness of the artificiality of the game world in which they are characters. These changes are reflected in the behaviour and voices of the characters. As the ‘hosts’ Maeve and Dolores achieve more sentience, their behaviour becomes more naturalistic, and their voices become more inflected, cynical and self-aware.

In these cinematic examples, there has been a trend for robot voices to change from highly artificial and in low modality towards more naturalistic ones. As science fiction narratives explored possible worlds in which machines and humans would become increasingly indistinguishable, voice assistant technologies in the early 21st century were developing higher modality voices and personae with personality.

Modality and personae in smart voice assistants

Smart voice assistants are being developed in the context of increasingly rich robot personalities in popular culture, and increasingly sophisticated speech synthesis. In 2017, Amazon spent US$22.6 billion and Alphabet spent US$16.2 billion overall on research and development, of which a significant proportion was dedicated to smart voice assistants (Loeb, 2018). The big tech companies adapted the voices of female voice actors as the default voices for their devices. This choice was strategic, leveraging the social connotations of the feminine voice to produce voice and personae that would appear naturally sympathetic and helpful.

Apple used the Susan Bennett in designing Siri when it was introduced for the iPhone in 2010 (Kleinman, 2017). Microsoft used Jennifer Lee Taylor’s voice when it released Cortana on the Windows phone in 2014. Amazon launched the voice of Alexa on the Amazon Echo smart speaker in 2015. Google’s voice assistant, based on an unknown actor, was code-named Holly before it was introduced in 2016. In the same year, Google DeepMind announced a technology called WaveNet that used a form of artificial intelligence called a ‘deep convolutional neural network’ that allowed it to generate more natural-sounding voices (Van den Oord and Dieleman, 2016). Human–computer interaction researchers Cohen et al. (2016) argue that companies seem to have decided that ‘there should be one assistant who acts for the user across situations and devices' (p. 1034). She is helpful without being obsequious, warm without being sexual, intelligent without being arrogant.

The first principle in Google's Conversation Design guide for third-party developers is 'Give your VUI a personality' (Google, 2019). The personae were designed to display localised cultural competence in every home. In the Australian version, the Google Assistant knows about pavlova and galahs, and uses Australian slang expressions. It is important that the persona should have a gentle sense of humour. The assistants make light of any dystopian associations. When asked 'Alexa, are you dangerous?', she replies in a calm retort, 'No, I am not dangerous'. When asked if she is Skynet from Terminator, Google Assistant sometimes says: 'I'm glad I'm not. It's more focussed on extermination than helpfulness. Skynet would make a terrible Google assistant'. These are just a few of the many anecdotes that were collected as part of our ongoing research on smart voice assistants, having introduced these into our homes and observed our own and our family members' reactions to these over an extended period.

A key design feature for voice assistants is that they should only speak once they are spoken to. The user must perform the trigger phrase, known as the wake-word, to invoke the assistant's attention. The companies assure users that their voice is recorded only after the wake-word is detected. This is not merely a technicality but part of the social design to create intelligible and unthreatening female voices that would also guard against the potential for these devices to expose the home to hackers, corporations or governments. The public relations sensitivity of this set-up was highlighted in the news coverage that followed when some users reported that Alexa had been heard laughing, without being initiated by the wake-word (Chokshi, 2018; Liao, 2018). There was another controversy when some Google Home Mini devices were found to have faulty switches that made them record without being called upon (Charlton, 2017). Google responded quickly by disabling the switch altogether (Russakovskii, 2017). There have been several reports in the news of incidents in which voice assistants have been triggered inadvertently, and even sent messages to strangers (Kim, 2018). These controversies highlight the cultural anxieties and tensions that surface when the wake-word proves fallible, and the faithful assistant betrays its promised confidentiality.

Invocationary acts and voice assistants

Another defining feature in the design of voice assistants with agency and personality is the way these mediate interactions using calls and responses. An interaction with a smart speaker is usefully seen as a form of invocation. Developers refer to it as such (Google, 2018). The device is silent until a user speaks the wake-word, followed by a question, command or other invocation. The smart voice assistant immediately responds with its own voice. This structure of mediation recalls an ancient form of speech act that uses the voice to call upon immediate guidance or assistance at a moment of crisis (Chesher, 2004). Invocations have been favourites of Homeric heroes and romantic poets, but with computers they are translated into digital form. In this modern context, invocations are still performed, even if the crisis is more mundane, such as a desire for music, a weather report or to turn on the light. The highly tailored response responds as if by magic without revealing the mechanisms by which the answer is constructed.

Voice assistants mobilise technologies with natural language capabilities and artificial intelligence (Dasgupta, 2018). Alexa, Siri and the Google Assistant recognise speech, interpret its meaning and pass messages back. However, as Austin (1962) and Searle (1969) systematically observed, everyday language does not just carry meanings ('constatives'), it actually does things ('performatives'). For example, we use language to make promises, give warnings or ask questions. We perform speech acts by making certain utterances ('I promise' or 'I warn'). If these speech acts are 'felicitous', using Austin's (1962) terminology, the listener will recognise this act and respond accordingly (p. 22). If the speech acts are not felicitous, it is not because they are factually wrong, but because they failed to produce the intended effect and did not lead to desired outcomes. The ability for voice assistants to respond to users' invocations is an extension of everyday natural speech and can be understood as an invocationary act.

We would suggest that the common question of whether smart speakers are 'intelligent' over-emphasises cognition and underplays their capacity to receive and perform everyday speech acts. While Alexa and the Google Assistant do not pass as intelligent for long, they do succeed in giving users' an experience of engaging with a conversational agent (Chen et al., 2019). In receiving invocationary speech acts, they translate speech reasonably accurately (automatic speech recognition), and respond to the users' invocations. If asked to set a timer they will, as promised, sound an alarm after a set time. In giving apparently sensible responses, they perform their own speech acts. For example, if they give a weather report of rain, and the user chooses to bring an umbrella. In each case, the smart speaker mediates invocationary speech acts, and influences users to take actions that might otherwise not be taken at that time.

Google's documentation (2018) requires that invocations must have an 'intent' – the goal or task that users want to do. Invocations can be made with different phrasings within certain programmed thresholds. Not all invocations work. Austin refers to failed everyday speech acts as 'misfires' (p. 16), such as when a promise is not made, or a request is not answered. There are equivalents in invocations to voice assistants. Saying the wake-word might not trigger the device, in which case the invocation fails, and nothing happens. In another situation, the wake-word might work, but the device might not make sense of your invocation, in which case, the invocation fails in a different way, and the assistant might respond with something like: 'My apologies, I don't understand'. In another case, your invocation might get a response, but not what you think you asked for. Note that these examples resemble misfires in face-to-face conversation, but with smart voice assistants, misfires can also be technical breakdowns in the device. Certain groups of users – those with strong accents, people with speech impediments, the very old or the very young – are more likely to have misfires because their speech is not within parameters of comprehensibility programmed into the smart algorithms. These groups may find these devices useless, as their speech falls outside the bounds of the 'normal'. On the other hand, people with visual impairment may find them assistive (Kalish, 2018).

As these voices are owned by large corporations, these invocational platforms risk reproducing social biases and compromising media pluralism and diversity. The assistants give only one brief answer to each question and draw these responses from a small range of sources. This gives the companies significant 'soft power' in their potential to influence feelings, thoughts and behaviour (Nye, 2004). It is notable that certain forms of knowledge are deemed too controversial to be answered. For example, in September 2018, asking the Google Assistant the questions ‘Who is Jesus?’, ‘Who is Buddha?’ and ‘Who is the devil?’ were all answered with the response: ‘Religion can be complicated, and I am still learning’. The responses of voice assistants can create their own controversies. When Apple’s Siri was first released in 2011 and users experimented with commands such as ‘Please go fuck yourself’ and ‘Hey Siri, you’re a bitch’, she responded with the inappropriately submissive: ‘I’d blush if I could’. The artificial intelligence has since been updated to reply: ‘I don’t know how to respond to that’ (Bergen, 2016; Newman, 2018). These issues evoke long-standing ethical and policy concerns of media power and influence, which need also be considered in relation to voice assistants. This is particularly so given they are sold as friendly helpers around the home rather than new media.

Pre-domestication and machinic orality

In her study of Google Glass, Saariketo (2018), building on Silverstone and Haddon’s (1996, 1998) contributions, suggested that ‘pre-domestication’ is an important but neglected phase of consumption in which ‘people need to be attracted, invited, and interpellated to familiarise [sic] with new technology as its potential future users’ (p. 1). While domestication explains the process by which consumers customise mass-produced products after buying them, pre-domestication refers to what Silverstone and Haddon (1996) call ‘an anticipation in design itself of the artefact’s likely place in (in this case) the home’ (p. 51). Saariketo focuses on the media coverage of Google Glass in the Finish Press in their beta phase to argue that this is one of the main ways that potential consumers and the public at large are influenced and familiarised with new media before their domestication and appropriation in users’ everyday lives.

In addition to the physical design of a device and its media representations, which have been the focus of pre-domestication before, we argue that the design of the voice in smart technologies is also a way the meanings of products are articulated. Smart voice assistants are pre-domesticated with their pre-packaged default voices, positioning them in the role of gendered helpers, with voices that are apparently natural, transparent and depoliticised. Through practices in voice design using *high modality*, a *warm and smart persona* with the programmed capacity to respond to *invocationary acts*, consumers’ relationships with their smart speakers have been prepared, even before they adopt them, in the form of the non-threatening voice of the female assistant.

In portending the future of media, Guattari (1995) identified a world mediated by machinic orality, bringing about what he described as

> . . . the junction of informatics, telematics, and the audiovisual will perhaps allow a decisive step to be made in the direction of interactivity, towards a post-media era and, correlatively, an acceleration of the machinic return of orality. The era of the digital keyboard will soon be over; it is through speech that dialogue with machines will be initiated . . . (p. 97)

Guattari’s 1990s optimism did not anticipate the extent to which voice assistants would be imbricated in corporate-dominated media ecologies and surveillant capitalist assemblages (Zuboff, 2018). Nevertheless, we contend there is value in his concept of ‘machinic orality’ because it gestures towards a regime of communicative power based on technologised voice interaction. Guattari’s (1995) conception of the machinic encompasses a degree of emergent indeterminacy that exceeds technical operation. Voice-controlled technologies manifest some of the dynamic, fluid and enunciative principles of ‘machinic orality’, a concept Guattari proposed against structuralist tendencies of perceiving meaning to be fixed in media form (Hetrigck, 2014). Nevertheless, the apparent naturalness of smart speakers belies the way in which these voices are pre-structured by cultural and technical processes.

Voice assistants offer a seductive power of the quasi-magical invocation with the potential to transform and guide everyday actions. Indeed, part of the success of these devices in the market is in their ability to get things done: ‘voice is proving a quick and convenient way of managing a range of tasks’ (Newman, 2018: 34). With the consolidation of smart voice media into the hands of a small number of global high-tech firms, the performative power of smart voice assistants may turn out to be not so ‘soft’ after all, and represent a new front for bias, hegemony, misinformation, fake news and other forms of media power. Changing media practices as a result of the use of these devices has already been observed with 65% of device owners listening to more music and consuming more news, podcasts and audiobooks (Newman, 2018).

The concept of machinic orality also locates the voice interaction paradigm within previous histories of media. As we have argued, the design of the voice in the current generation of smart voice assistants is informed by robotic voices in popular culture and scientific discourse since the early 20th century, and indeed by the ancient dramaturgical traditions of vocal performance (Ihde, 2007). VUIs echo previous practices in media that established a voice of authority, such as Roosevelt’s fireside chats, DJs in radio and the news anchor in television. In this sense, we suggest smart speakers are a medium for creating on-demand voices that similarly reach into domestic spaces. However, with its still artificial cosmopolitan voice and its unreliable handling of different accents, its localisation for different places is far from fully realised.

Furthermore, machinic orality keys smart voice assistants into global vectors that ‘mediate the relationship users and potential users have with late-capitalist market logics in the platform economy’ (Woods, 2018: 1). Voice assistants have the capacity to trace the ‘behavioural surplus’ of every user invocation according to the logics of surveillance capitalism (Zuboff, 2018: 102). The invoker is invoked. Just as Google traces users’ activities across all its platforms, and Amazon tracks every mouse click on its site, smart voice assistants are intimate domestic witnesses or even spies (Fowler, 2019). Both Amazon and Google ameliorate this relationship by giving users access to their conversation history, and the capacity to manage it. They also use that history to track users’ tastes and interests, and even listen to personal voice recordings, supposedly to improve speech performance. As e-commerce devices, smart speakers mobilise strategies that channel consumption through default monopolistic pathways. They summon information drawn from the knowledge repositories of their commercial partners, for example, in announcing news bulletins and answering users’ questions, thus authorising certain forms of knowledge and displacing other media.

As Woods (2018) has argued, and on which we concur, the feminine Muse-like persona is strategically leveraged to articulate positive meanings and steer would-be consumers away from thinking of them as dangerous surveillant machines. As we have shown, the professional feminine voice also serves to instate a vernacular authority that distances itself from negative associations of synthesised and processed voices in cinematic depictions of the menacing (male) or excessively overprotective (mother) robot or artificial intelligence. Modal distinctions in dialogue, performance and sound manipulation are similarly used to consolidate the perception of a natural human voice. Voice technology and the use of a naturalistic persona supported by artificial intelligence appear to make the medium itself invisible (Phan, 2017). Just as Crawford and Joler (2018) observe, the apparent simplicity of the user–assistant interaction belies the massive complexity of the extended human, informational, material and energy networks that enable these apparently trivial encounters.

There are moves towards making the interactions with assistants more complex by including multiple voices and adding the ability to carry out ‘continued conversation’, and even to allow the assistant to initiate dialogue. Google’s experimental technology Duplex allows users to ask the assistant to make phone calls on their behalf to perform tasks such as booking a hair appointment. In this case, the assistant starts a naturalistic conversation with an involuntary user. The test of its success here is the extent to which it/she can pass as ‘human’. These developments further risk manipulating consumers through the technocultural construction of the voice of the smart assistant and obscure the implications of surveillance, soft power and global monopoly.

Conclusion

This article has added to critical literature on the design of female personae and voices in smart voice assistants by highlighting how the cultural imaginary functions as an arena for negotiating the meanings of the robotic voice against pre-existing social connotations of gender, class and race. We have argued that the emerging discipline of VUI design applied in the current generation of smart voice assistants has been informed by the conventions of robotic voices established in 20th- and 21st-century screen culture and scientific discourse. In making this connection, we are showing the role of cultural legacies in the emergence of new media innovations and the specific selections and exclusions made by manufacturers in design processes. We explored a number of tropes in cinematic and televisual robots and introduced the concepts of modality, personae and invocationary acts to help explain the material features of the robotic voice used by designers to construct the voice of smart speakers. We identified processes of ‘pre-domestication’ to explain the way that smart voice assistants are prepared for consumers and adapted ‘machinic orality’ to capture an emerging regime of power based on technologised voice interaction. Our investigation points to the need for further research to better understand the role of voice interaction as a new medium of power and influence in consumption practices.

What began with the voice is now being pursued with the simulation of facial expressions and choreographed actorly gestures to allow robots and humans to interact with meaning and emotion (Zhao, 2006). The smart speaker illustrates how a new media form can perform speech acts and become imaginable as a social actor. However, this actor is animated by corporations within a context of surveillance capitalism. We need new conceptual tools and frameworks to critically interrogate their implications and understand these new developments.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The authors received funding from the School of Literature, Art and Media at the University of Sydney for research assistance.

ORCID iDs

Justine Humphry https://orcid.org/0000-0002-2376-2089

Chris Chesher https://orcid.org/0000-0001-9377-4512

References

Austin JL (1962) *How to Do Things with Words*. Oxford: Clarendon Press.

Bergen H (2016) ‘I’d blush if I could’: digital assistants, disembodied cyborgs and the problem of gender. *Word and Text VI*: 95–113.

Bishop A (2014) Android problems: the representation of robots in cinema. *The Artifice*. Available at: https://the-artifice.com/representation-robots-cinema/ (accessed 29 July 2019).

Bucholtz M and Hall K (2009) Locating identity in language. In: Llamas C and Watt D (eds) *Language and Identities*. Edinburgh: Edinburgh University Press, pp. 18–28.

Čapek K, Playfair N and Selver P (1928) *R.U.R. (Rossum’s Universal Robots): A Play in Three Acts and an Epilogue*. London: Humphrey Milford; Oxford University Press.

Charlton C (2017) Google Home Mini caught constantly spying in owner’s bathroom. *Gearbrain*, 11 October. Available at: https://www.gearbrain.com/google-home-mini-recorded-everything-2495445625.html (accessed 31 October 2019).

Chen X, Mi J, Jia M, et al. (2019, October) Chat with smart conversational agents: how to evaluate chat experience in smart home. In: *Proceedings of the MobileHCI’19: 21st international conference on human-computer interaction with mobile devices and services*, Taipei, 1–4 October, pp. 1–6. ACM. Available at: https://dl.acm.org/doi/pdf/10.1145/3338286.3344408 (accessed 7 April 2020).

Chesher C (2001) *Computers as invocational media*. PhD Thesis, Macquarie University, Sydney, NSW, Australia.

Chesher C (2004) Hyperlink as invocationary act. *Australia and New Zealand communications association conference*, Sydney, 7–9 July.

Chokshi N (2018) Amazon knows why Alexa was laughing at its customers. *New York Times Online*. Available at: https://www.nytimes.com/2018/03/08/business/alexa-laugh-amazon-echo.html (accessed 31 October 2019).

Cohen P, Cheye A, Horvitz E, et al. (2016) On the future of digital assistants. *CHI’16 Extended Abstracts*, 7–12 May, San Jose, CA. Available at: http://dx.doi.org/10.1145/2851581.2886425

Cox TJ (2018) *Now You’re Talking: Human Conversation from the Neanderthals to Artificial Intelligence*. London: The Bodley Head.

Crawford K and Joler V (2018) *Anatomy of an AI system*. New York: AI Now Institute and Share Lab.

Creed B (1993) *The Monstrous-Feminine: Film, Feminism, Psychoanalysis*. London; New York: Routledge.

Cutsinger and Akersh (2018) How building for voice differs from building for screen. Alexa. design/webmobiletovoice. *Amazon Alexa*. Available at: https://register.gotowebinar.com/recording/3829485040839408386 (accessed 1 October 2018).

Dasgupta R (2018) *Voice User Interface Design: Moving from GUI to Mixed Modal Interaction*. New York: Apress.

Davis A (2009) Investigating cultural producers. In: Pickering M (ed.) *Research Methods for Cultural Studies*. Edinburgh: Edinburgh University Press, pp. 53–67.

Eschner K (2017) Meet Pedro the 'Voder', the first electronic machine to talk. Available at: https://www.smithsonianmag.com/smart-news/meet-pedro-voder-first-electronic-machine-talk-180963516/#ilZpa2Jk4FkuplmZ.99 (accessed 1 October 2018).

Fowler GA (2019) Alexa has been eavesdropping on you the whole time. *The Washington Post*, 6 May. Available at: https://www.washingtonpost.com/technology/2019/05/06/alexa-has-been-eavesdropping-you-this-whole-time/

Google (2018) Conversation Design. Available at: https://developers.google.com/assistant/actions/design

Google (2019) Create a persona. *Conversation Design*. Available at: https://designguidelines.withgoogle.com/conversation/conversation-design-process/create-a-persona.html

Goulden M (2019) 'Delete the family': platform families and the colonisation of the smart home. *Information, Communication and Society*. Available at: https://www.tandfonline.com/doi/full/10.1080/1369118X.2019.1668454 (accessed 23 October 2019).

Guattari F (1992) Machinic orality and virtual ecology. In: Trans. Bains P and Pefanis J (eds) *Chaosmosis: An Ethico-Aesthetic Paradigm*. Bloomington: Indiana University Press, pp. 88–97.

Guattari F (1995) *Chaosmosis: An Ethico-Aesthetic Paradigm*. Bloomington: Indiana University Press.

Hetrigck J (2014) Video assemblages: 'Machinic animism' and 'asignifying semiotics' in the work of Melitopoulos and Lazzarato. *Footprint* 8: 53–68.

Hui JY and Leong D (2017) The era of ubiquitous listening: living in a world of speech-activated devices. *Asian Journal of Public Affairs* 10(1): e5.

Ihde D (2007) *Listening and Voice: Phenomenologies of Sound*. 2nd ed. Albany, NY: State University of New York Press.

Kalish J (2018) Why Amazon's Alexa is 'life changing' for the blind. *PC*. Available at: https://au.pcmag.com/news/51189/why-amazons-alexa-is-life-changing-for-the-blind (accessed 21 July 2019).

Katz M (2010) *Capturing Sound: How Technology Has Changed Music*. Berkeley, CA: University of California Press.

Kim E (2018) Amazon Echo secretly recorded a family's conversation and sent it to a random person on their contact list. *CNBC*. Available at: https://www.cnbc.com/2018/05/24/amazon-echo-recorded-conversation-sent-to-random-person-report.html (accessed 31 October 2019).

Kinsella B (2018) Smart speakers to reach 100 million installed base worldwide in 2018, Google to Catch Amazon by 2022. *Voicebot.ai*. Available at: https://voicebot.ai/2018/07/10/smart-speakers-to-reach-100-million-installed-base-worldwide-in-2018-google-to-catch-amazon-by-2022/ (accessed 1 October 2018).

Kleinman A (2017) Meet the woman who says she's the voice of Siri. *The Huffington Post*. Available at: https://www.huffingtonpost.com.au/entry/voice-siri_n_4043134 (accessed 1 May 2019).

Krapp P (2011) *Noise Channels Glitch and Error in Digital Culture*. Minneapolis, MN: University of Minnesota Press.

Kriz S, Ferro TD, Damera P, et al. (2010) Fictional robots as a data source in HRI research: exploring the link between science fiction and interactional expectations'. In: *19th international symposium in robot and human interactive communication*, Viareggio, 13–15 September, pp. 458–463. New York: IEEE.

Lanser SS (1992) *Fictions of Authority: Women Writers and Narrative Voice*. Ithica, NY: Cornell University Press.

Liao S (2018) Amazon has a fix for Alexa's creepy laughs. *Circuit Breaker*, 7 May. Available at: https://www.theverge.com/circuitbreaker/2018/3/7/17092334/amazon-alexa-devices-strange-laughter (accessed 16 October 2019).

Linngard R (1985) *Electronic Synthesis of Speech*. Cambridge: Cambridge University Press.

Loeb W (2018) Amazon is the biggest investor in the future, spends $22.6 billion on R&D. *Forbes*. Available at: https://www.forbes.com/sites/walterloeb/2018/11/01/amazon-is-biggest-investor-for-the-future/#cf811411f1db (accessed 14 October 2019).

Mortensen D (2019) How to design voice user interfaces. *Interaction Design Foundation*. Available at: https://www.interaction-design.org/literature/article/how-to-design-voice-user-interfaces (accessed 25 July 2019).

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

Newman N (2018) Journalism, media, and technology trends and predictions 2018. *Digital News Report* 2018. Reuters Institute for the Study of Journalism; University of Oxford, Oxford.

Nielsen L (2013) *Personas-User Focused Design*. London: Springer, pp. 59–79.

Nye JS (2004) *Soft Power: The Means to Success in World Politics*. New York: Public Affairs.

Ong W (2002[1982]) *Orality and Literacy: The Technologizing of the Word*. New York: Routledge.

Phan T (2017) The materiality of the digital and the gendered voice of Siri. *Transformations*, 29 Issue. Available at: http://www.transformationsjournal.org (accessed 1 October 2018).

Phan T (2019) Amazon Echo and the aesthetics of whiteness. *Catalyst: Feminism, Theory, Technoscience* 5(1): 1–38.

Pieraccini R (2012) *The Voice in the Machine: Building Computers That Understand Speech*. Cambridge, MA: MIT Press.

Roemmele B (2016) The Voder: 1939, the worlds first electronic voice synthesizer. Available at: https://youtu.be/TsdOej_nC1M (accessed 23 October 2019).

Roettgers J (2019) How Google found its voice. *Variety*, 18 September. Available at: https://variety.com/2019/digital/features/google-assistant-name-personality-voice-technology-design-1203340223/ (accessed October 23 2019).

Russakovskii A (2017) Google is permanently nerfing all Home Minis because mine spied on everything I said 24/7 [Update x2] Android police. Available at: https://www.androidpolice.com/2017/10/10/google-nerfing-home-minis-mine-spied-everything-said-247/ (accessed 16 October 2019).

Rydell RW (1990) Selling the world of tomorrow: New York's 1939 World's Fair. *The Journal of American History* 77(3): 966–970.

Saariketo M (2018) The unchallenged persuasions of mobile media technology: the pre-domestication of Google Glass in the Finnish press. In: *Proceedings of the digital humanities in the Nordic countries 3rd conference* (eds E Mäkelä, M Tolonen and J Tuominen), Helsinki, 7–9 March. CEUR Workshop Proceedings, Vol. 2084, pp. 454–459. Helsinki: Digital Humanities in the Nordic Countries.

Sadoski M (1992) Imagination, cognition, and persona. *Rhetoric Review* 10(2): 266–278.

Schafer RM (1969) *The New Soundscape: A Handbook for the Modern Music Teacher*. Vancouver, BC, Canada: BMI Canada.

Searle J (1969) *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

Silverstone R and Haddon L (1996) Design and the domestication of information and communications technologies: technical change and everyday life'. In: Mansell R and Silverstone R (eds) *Communication by Design: The Politics of Communication Technologies*. Oxford: Oxford University Press, pp. 44–74.

Silverstone R and Haddon L (1998) The domestication of ICTs: households, families, and technical change. In: Mansell R and Silverstone R (eds) *Communication by Design: The Politics of Information and Communication Technologies*. Oxford: Oxford University Press, pp. 44–74.

Smith J (2008) Tearing speech to pieces: voice technologies of the 1940s. *Music, Sound, and the Moving Image* 2(2): 183–206.

Sofoulis Z (2001) Smart spaces@ the final frontier. In: Munt SR (ed.) *Technospaces: Inside the New Media*. London: Continuum, pp. 129–146.

Statista Research Department (2020) Number of digital voice assistants in use worldwide from 2019 to 2023 (in billions)* Statista (website). Available at: https://www.statista.com/statistics/973815/worldwide-digital-voice-assistant-in-use/ (accessed 9 November 2019).

Strengers Y and Nicholls L (2017) Aesthetic pleasures and gendered tech-work in the 21st-century smart home. *Media International Australia* 166(1): 70–80.

Uotinen J (2010) Digital television and the machine that goes 'PING!': autoethnography as a method for cultural studies of technology. *Journal for Cultural Research* 14(2): 161–175.

Van den Oord A and Dieleman S (2016) WaveNet: a generative model for raw audio. Available at: https://deepmind.com/blog/article/wavenet-generative-model-raw-audio (accessed 16 October 2019).

Van Leeuwen T (1999) *Speech, Music, Sound*. Basingstoke: MacMillan Press.

West M, Kraut R and Chew HE (2019) *'I'd Blush If I Could'. Closing Gender Divides in Digital Skills through Education*. UNESCO Report, EQUALS, Paris.

Woods HS (2018) Asking more of Siri and Alexa: feminine persona in service of surveillance capitalism. *Critical Studies in Media Communication* 35(4): 334–349.

Zhao S (2006) Humanoid social robots as a medium of communication. *New Media & Society* 8(3): 401–419.

Zuboff S (2018) *Surveillance Capitalism*. London: Profile Books.

Author biographies

Justine Humphry is a lecturer in Digital Cultures in the Department of Media and Communications at the University of Sydney. Her research is on the cultures and politics of mobile media and smart technologies in everyday life.

Chris Chesher is a senior lecturer in Digital Cultures in the Department of Media and Communications at the University of Sydney. His current research concerns smart technologies in the home and the city, social robots, and invocational media.