VOICES ECHOING IN THE CLOUDS

ECOS DE VOZES NAS NUVENS

Manuel Portela
Centro de Literatura Portuguesa
Universidade de Coimbra

ABSTRACT

Natural language technologies are now able to listen to and process our voices and languages in real time. Voice-controlled digital assistants have emerged as multipurpose human-machine interfaces. We can train them to recognize our particular speech patterns and we can talk to them. Duplex and Alexa – two voice-controlled cloud-computing services – are described as instances of the datafication of language and subjectivity, and as archaeological echoes of telephonic technologies. The phantasmatic and acousmatic resonance of a disembodied and ubiquitous voice is the ultimate aural-oral embodiment of the human. Through algorithmic transactions between listening and speaking, the naturalization of computer-mediated communication obscures the deep commodification of symbolic exchange. At the same time, voice and language are revealed as technologies of the human.

Keywords: voice assistant, cloud computing, artificial speech, grammalepsis, aurature, John Cayley

RESUMO

As tecnologias de linguagem natural são agora capazes de escutar e produzir fala em tempo real. Os assistentes digitais ativados por voz tornaram-se

https://doi.org/10.14195/2183-847X_10_39

interfaces multifuncionais. Podemos treiná-los de modo a reconhecerem vozes e prosódias específicas e podemos conversar com eles e elas. Duplex e Alexa – dois serviços de computação em nuvem controlados por voz – são descritos como instâncias da informatização da linguagem e da subjetividade, e como ecos arqueológicos das tecnologias telefónicas. A ressonância fantasmática e acusmática de uma voz desencarnada e ubíqua torna-se a derradeira personificação aural-oral do humano. Através de transações algorítmicas entre ouvir e falar, a naturalização da rede e da comunicação mediada por computador obscurece a profunda mercantilização da troca simbólica. Ao mesmo tempo, voz e língua revelam-se como tecnologias do humano.

Palavras-chave: assistente de voz, computação em nuvem, fala artificial, gramalepsia, auratura; John Cayley

> A long-standing goal of human-computer interaction has been to enable people to have a natural conversation with computers, as they would with each other. (Leviathan and Matias, 2018)

1. INTRODUCTION

In 1965, Ivan Sutherland described the “ultimate display” in the following terms:

> The ultimate display would, of course, be a room within which the computer can control the existence of matter. A chair displayed in such a room would be good enough to sit in. Handcuffs displayed in such a room would be confining, and a bullet displayed in such a room would be fatal. With appropriate programming such a display could literally be the Wonderland into which Alice walked. (Sutherland 1965)

The sheer mathematical force of bringing matter into existence was conceived by Sutherland as a visual dreamlike virtual reality, perfectly fitted to human vision and proprioception. This powerful kinesthetic display would have the ability "to serve as many senses as possible." Yet Sutherland is aware of the sound limitations of what computers could do in 1965: "Excellent audio displays exist, but unfortunately we have little ability to have the computer produce meaningful sounds." Fifty-five years later this situation has significantly changed: computers are able to produce and understand speech. Voice user interfaces have thus added an additional level of immersive complexity to kinesthetic displays by mixing human and computer-generated voices in what is increasingly advertised as a seamless integration of human and machine.

This article will address the ongoing transition from graphical user interfaces to voice user interfaces.[1] Natural language technologies are able to listen to and process our voices and languages in real time. Voice-controlled digital assistants have emerged as a multipurpose human-machine interface. We can train them to recognize our particular speech patterns and we can talk to them. Duplex – Google's artificial intelligence voice system – and Alexa – Amazon's voice-controlled cloud-based service – will be described as instances of the datafication of language and subjectivity, and as a legacy of earlier media technologies and social practices. The

[1] The first version of this text was presented on July 3, 2018, at the VIII Lisbon Summer School for the Study of Culture ("Cyber+Cipher+Culture", 2-7 July 2018), during the annual meeting of the Lisbon Consortium's MA and PhD Programme in Culture, Portuguese Catholic University. A second version was presented on June 14, 2019, at the XIII Jornadas de Investigação do CIAC – Centro de Investigação em Artes e Comunicação, University of Algarve. I want to thank the organizers of both events, particularly Isabel Capeloa Gil (Catholic University), Mirian Tavares (CIAC), and Sandra Boto (CIAC).

phantasmatic and acousmatic resonance of a disembodied generated voice is the aural-oral embodiment of the ultimate display. Through algorithmic transactions between listening and speaking, the naturalization of computer-mediated communication obscures the deep commodification of symbolic exchange.

2. DUPLEX

On May 8, 2018, Google announced its latest version of the technology behind Google Assistant. Called Google Duplex, the new system has been press released as "The future of the Google Assistant: Helping you get things done to give you time back" (Huffman 2018). Ease of use and production efficiency is the familiar mantra for constant software and hardware upgrades of our digital devices. Like similar voice assistants from other big data corporations, Google Duplex is part of an ongoing shift in human-computer interaction towards spoken natural language interfaces, which promises to increase the universality and transparency of interactions. The multidevice and multicultural scale of its availability is precisely the first highlight of Duplex's self-advertisement:

> As of today, the Google Assistant is available on more than 500 million devices, it works with over 5,000 connected home devices, it's available in cars from more than 40 brands, and it's built right into the latest devices, from the Active Edge in the Pixel 2 to a dedicated Assistant key in the LG G7 ThinQ. Plus, it'll be available in more than 30 languages and 80 countries by the end of the year. (Huffman 2018)

The second point is the fact that users can speak naturally to their devices through Google Assistant, and they can also choose from a menu of customized human-like voices – six new voices referred to below (Google Assistant 2018a). Technological advances in artificial intelligence voice-processing (Leviathan and Matias 2018) thus relate both to the aural and oral abilities of voice assistants to engage in natural conversation:

> We've dramatically improved our language understanding so you can speak naturally to your Google Assistant and it will know what you mean. [...] One of the most important parts of the Assistant is its voice–it needs to feel both personal and natural. Up until now, creating a new voice took hundreds of hours in a recording studio. But with advancements in AI and WaveNet technology from DeepMind, we can now create new voices in just a few weeks and are able to capture subtleties like pitch, pace, and all the pauses that convey meaning, so that voices are natural-sounding and unique. Soon you'll be able to have a natural back-and-forth conversation without repeating "Hey Google" for each follow-up request. The Assistant will be able to understand when you're talking to it versus someone else, and will respond accordingly. (Huffman 2018)

The ability of the Duplex Google Assistant to engage in conversation was dramatically staged at the company developer's May 8th demo by means of two recorded phone interactions involving the voice assistant and a hair salon, in one instance, and the voice assistant and a restaurant, in the other (Google Assistant 2018c; 2018b).[2] Articles written about the launch of the new assistant emphasized the strangeness of witnessing the human-like features of

[2] The recording of those interactions, which was broadcast during the demo, is available here: "Duplex scheduling a hair salon appointment" (May 8, 2018) http://www.gstatic.com/b-g/DMS03IIQXU3TY2FD6DLPLOMBBBJ2CH188143148.mp3; "Duplex calling a restaurant" (May 8, 2018): http://www.gstatic.com/b-g/KOK4HAMTAPH5Z96154F6GKUM74A3Z1576269077.mp3

the new synthetic voices, including their subtle domain of pause and intonation: “Google's robot assistant now makes eerily lifelike phone calls for you” was the title of the piece in *The Guardian* (Solon 2018); “Are Google's A.I.-Powered Phone Calls Cool, Creepy, or Both?,” asked *The New York Times* technology columnist (Roose 2018); “So, Umm, Google Duplex's Chatter Is Not Quite Human” was the ironic title chosen by the systems expert in the *Scientific American* blog (Greenemeier 2018b).

Aware of the ethical questions raised by fooling the person on the other side of the phone into believing that s/he is talking to an actual person, Google engineers make the case that the artificial context of the proxy speech actant will also become clear:

> Powered by a new technology we call Google Duplex, the Assistant can understand complex sentences, fast speech, and long remarks, so it can respond naturally in a phone conversation. Even though the calls will sound very natural, the Assistant will be clear about the intent of the call so businesses understand the context. (Leviathan and Matias 2018)

In Frank Nickel's extension of speech act theory to artificial speech, voice assistants are described as “proxy speech actants” (Nickel 2013), which means that they act on behalf of someone and thus embody the intentionality and legal responsibility inherent in the communication situation – in these instances, the assistant's speech acts were made on behalf of the persons making the haircut and the dinner appointments, respectively.

Current Artificial Intelligence methods for processing human speech – understood as the ability both to have a voice and to understand a voice – suggest that voice user interfaces will become far more significant in future interface design, perhaps more significant than current graphical user interfaces. Artificial voice interfaces have also been providing answers to three related questions that have fueled research by neuroscientists, linguists and software engineers: What is a human a voice and how do we recognize it? How are voice perception and voice production related? What enables a voice to acquire linguistic features? Latinus and Belin summarize the acoustic and linguistic features of voice as follows:

> (…) vocal sounds are generated by the interplay of a source (the vocal folds in the larynx) and a filter (the vocal tract above the larynx). The most common vocal sounds ('voiced sounds') correspond to a periodic oscillation of the vocal folds with a well-defined fundamental frequency (f0). The range of f0 values a given individual can achieve during normal phonation or singing is fairly extended, but the average f0 of an individual is largely a function of the size of the vocal folds: men have much larger vocal folds than women or children, resulting in generally lower f0 values. The vocal tract above the larynx acts as a filter reinforcing certain frequencies of the source, called 'formants'. Formant frequencies depend on the particular configuration of the articulators during speech, but also on the individual's vocal tract size. (...) Linguistic information is essentially conveyed by changes in formant frequencies. (Latinus and Belin 2011: 143)

From the point of view of production, the simulation of the acoustic features of human voices engaged in speech acts depends on the specific modeling of the relations between fundamental and formant frequencies, on the one hand, and on the control of subtle changes in formant frequencies from which a phonological system of differences emerges, on the other.

Along with the production of natural language, the understanding of the cognitive and affective perception of voice is also a major research topic in the domain of language technologies using artificial intelligence methods such as recurrent neural networks. Clifford Nass and Scott Brave, in their classic experiment-based work about voice interfaces in *Wired for Speech* (first published in 2005), argue that "listeners and talkers cannot suppress their natural responses to speech, regardless of source" (Nass and Brave 2005: 4), which means that technology-based voices are processed in the same way as human voices, both in terms of brain activation and behavioral response related to conscious and unconscious inferences about voices. Voice, regardless of its physical embodiment or origin, thus seems a strong candidate for the best artificial surrogate of the human.[3] Nass and Brave described the growing importance of the voice as input and output in human-machine interactions, anticipating the enormous potential of user voice interface as technology for capturing data based on the "fundamental means of human communication" (1):

> As a result of these automatic and unconscious social responses to voice technologies, the psychology of interface speech is the psychology of human speech: voice interfaces are intrinsically social interfaces. Designers must create voice interfaces for brains that are obsessed with extracting as much social information as possible from speech and with using that information to guide attitudes and behaviors.

[3] Spike Jonze's fictional exploration of a voice-based operating system in *Her* (2014) addresses important cognitive and affective dimensions of artificial assistants based on emulation of human voices. Pedro Serra describes this paradoxical disembodiment of the voice as an expression of the voice itself as an external body, an avatar-voice: "The digital voice is Samantha's body; deep down it is an inhuman voice, but in the strict sense that it is also the very human voice: it only exists in separation from the body – it becomes a body beyond the body –, an avatar-voice determined by technologies that structure and repeat it, becoming the paradoxical support of fiction which is that of an intimate consciousness capable of subjective action and intersubjective relationship." [*my translation*] (Serra 2015: 18).

> Because humans will respond socially to voice interfaces, designers can tap into the automatic and powerful responses elicited by all voices, whether of human or machine origin, to increase liking, trust, efficiency, learning, and even buying. (Nass and Brave 2005: 4)

It is important to notice the range of perlocutory effects of voice interfaces listed by Nass and Brake: "to increase liking, trust, efficiency, learning, and even buying." If we relate this to Austin's speech act theory, it suggests that a layer of the action of language is performed through non-semantic information provided by specific sound patterns that define any given statement. The prosodic features of the voice are thus part of the rhetorical structure that structures a conversation and defines the social transaction that bring speaker and listener into existence.

When Alan Turing, in his 1950 article "Computing Machine and Intelligence", asked the question "can machines think?", he conceived of the test for telling humans and machines apart based on the *imitation game* (Turing 1950: 433-434). In this game an interrogator (C) has to determine, through a series of questions, which of two subjects (A or B) is the man or the woman. They are placed in a room apart from the interrogator and they have to typewrite their answers (which are instantaneously tele-printed in the other room) so that their voices or their handwriting do not give away their gender. It is interesting that the inferential game for proving that machines can think is related to withholding one's gender and one's voice. So, in Turing's envisioned future when learning machines have brought about artificial intelligence, the role of the interrogator would be not to tell man and woman apart, but attempt to distinguish human from machine.

> We now ask the question, 'What will happen when a machine takes the part of A in this game?' Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, 'Can machines think?' (Turing 1950: 434)

Turing dismisses all metaphysical and theoretical objections in favor of a pragmatic approach: the ability to engage in symbolic interaction through meaningful conversation abstracted in written form thus becomes the basis for the so-called Turing's test. The human-machine conversation mediated through the writing keyboard is also the basis for CAPTCHAS – the test for screening out bots from signing in into websites and filling in forms in a world of mostly written input.

Advances in voice processing over the last two decades have complicated Turing's test in the sense that engaging in a voiced conversation, however limited, may be said to provide an aural-oral version of the Turing test, that is, the test that determines whether a machine may or may not be distinguishable from a human. Telecommunication, as in Duplex's phone interactions mentioned above, is still a prerequisite – because of the speakers' homogenizing effect in hiding the corporeal identity of voice sources –, but the actual sensory perception of a female or male voice has become part of the post-Turing imitation game. The femaleness or the maleness of the voices thus become additional markers of the machine 's humanness, as witnessed by the six new Google Assistant voices synthesized at the moment (Google Assistant 2018a).[4] We may say that telling

[4] Those six artificial human-like voices are demonstrated in the following advertisement: "Google Assistant: Available in 6 new voices" (May 8, 2018) https://youtube.com/watch?v=3YQJOHqkhgM

female and male apart, that is the speech processor's ability to gender the voice, has become part of the imitation game for remixing human and machine through the acts of speaking and listening.

Available in 6 new voices

Figure 1. "Google Assistante: Available in 6 new voices" (May 8, 2018) [screen capture].

3. ALEXA

> Alexa is Amazon's voice service and the brain behind tens of millions of devices like the Amazon Echo, Echo Dot, and Echo Show. Alexa provides capabilities, or skills, that enable customers to create a more personalized experience. There are now more than 25,000 skills from companies like Starbucks, Uber, and Capital One as well as other innovative designers and developers. https://www.youtube.com/watch?v=UOEIH2l9z7c

Besides the usual emphasis on the naturalness of Alexa's enabled voice-based interactions, the above quoted description highlights its intimate relation with the internet's commercial infrastructure, particularly with large multinational corporations and their range of services. Alexa's set of skills is referred to as "infinite abilities" and Alexa-enabled devices such as Echo are presented as one more household appliance useful for a large range of needs and a great variety of people. The verbal and visual rhetoric of the following videos, dated from 2015 and 2016 (Amazon 2015; 2016), are clear about the ongoing naturalization of voice user interfaces as conversational partners.[5]

FIGURE 2. "What Is Alexa? An Introduction to Amazon's Alexa Voice Service" (September 14, 2016) [screen capture].

[5] The first one is addressed to software developers – "What Is Alexa? An Introduction to Amazon's Alexa Voice Service" (September 14, 2016, 1:34; https://www.youtube.com/watch?v=UOEIH2l9z7c) – while the second is aimed at general users – "Amazon Echo – Now Available" (June 23, 2015, 2:53; https://www.youtube.com/watch?v=FQn6aFQwBQU)

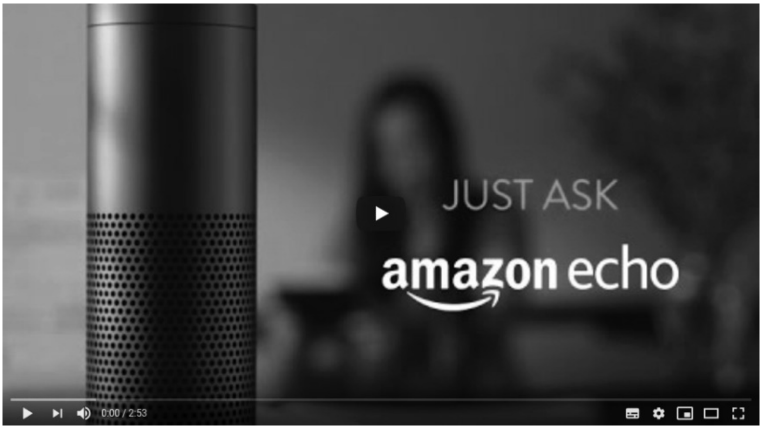

FIGURE 3. “Amazon Echo – Now Available” (June 23, 2015) [screen capture].

There are two historical legacies to bear in mind when thinking about the sweet, feminized synthetic voice of artificial intelligence avatars or the familiar human voice recordings for telephone recorded voice assistants or public announcement speakers, such as teller machines, gas stations, train stations, or airports, for instance. One is related to the feminization of women, and the other with the feminization of technology.

On the one hand, we have to look at the historical process (from the eighteenth and nineteenth-century onwards) of constructing a normative feminine identity, which came to associate a certain constellation of attributes – such as sweetness, sereneness, and caring – with women’s body language, including the voice of women. This feminization of women subjects, including the feminization of female voices, accompanies their passage from the private domestic sphere into the public professional sphere, which in the case of the middle class included schools and hospitals as transitional institutions:

teaching and nursing were two ways of legitimating the act of leaving the domestic sphere, producing an identity of the professional woman as an extension of the mother and her role as protector and guardian. Working-class women were moving from the countryside into the factories in the same period, but also finding employment as domestic servants in middle and upper-class homes where a similar normative code of feminine identity applied.

On the other hand, the history of media shows how the voice of auditory media also became extremely feminized, ever since the beginning of telephone networks in the late nineteenth century. After an initial period of experimentation with young male telephone operators, telephone corporations such as American Telephone & Telegraph almost universally turned to young women only, establishing strict codes of conduct that trained operators not only in what they should and were allowed to say but also trained them to produce certain patterns of vocal prosody. Millions of young women were trained as telephone operators throughout the 20th century. Until the establishment of automatic telephone networks in the 1950s and 1960s, the voice of the telephone was fully established as the feminized voice we continue to listen to today in call centers, recorded assistance, and public place announcements. It is a version of this feminized telephonic media voice that engineers are currently programming in synthetic speech processors and in cloud-based artificial intelligence voice-controlled interfaces, such as Alexa and Echo (Greenemeier 2018a).

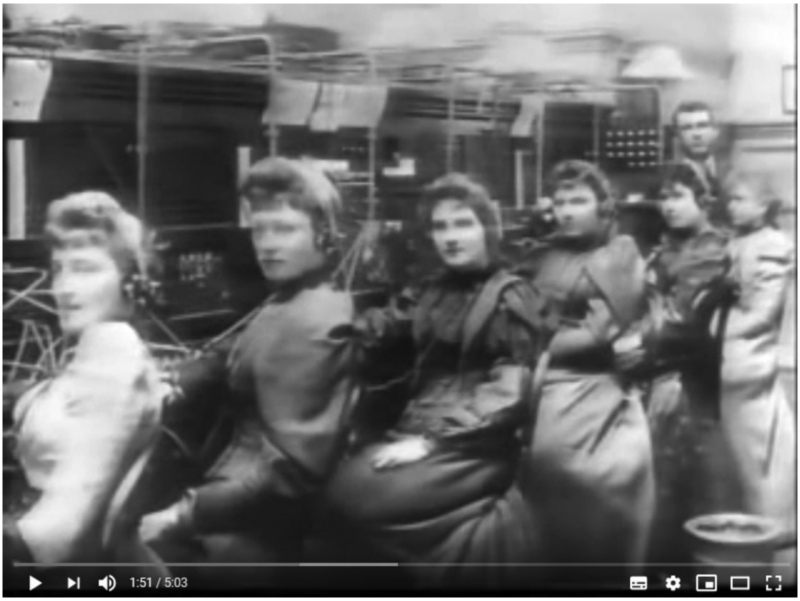

FIGURE 4. American Telephone and Telegraph Archives: “Women Telephone Operators” 1930s [screen capture].

One could argue that, from the perspective both of women’s social history and of media archaeology, Alexa’s voice is deeply integrated into the cultural unconscious and into the collective psyche. Therefore *its* – and my choice of personal pronoun is deliberate here – transparency as a natural talking and listening interface resides not only in *its* machinic efficiency but in the feminized prosody of *her* timbre and *her* phonetic articulation. It is a kind of naturalized acoustic hallucination (that is, a realized fantasy) that which allows a voice without body to “speak with” us or “speak to” us, and, even more significantly, to “listen to” us. Our “master’s voice” is also our “mother’s voice”, and both have also become our “master’s ear” and our “mother’s ear”. This is the point where we enter a through-thelooking-glass acoustic CAPTCHA, one in which we prove to be human because we are able to be listened to (not because we decipher alphanumeric characters or check a dialogue box stating that we are not a robot.) The hearing-speaking voice of Alexa opens up "the Wonderland into which Alice walked".

Writing has been the traditional modality of human input to the computer's processing and memory units insofar as keyboards became the general interface between software and hardware, between human code and machine code. Input mechanisms were extended to include light pen, computer mouse, touchscreen, microphone, camera, and other kinds of sensors. The expansion of the internet led to the integration of personal computers into a global network of servers and datacenters that process all network activity in real-time. As internet communication expanded and many kinds of human activity gained online expression, the acts of writing and reading became the major source for capturing user behavior. The acts of communication and symbolic consumption through connected cloud-based applications turned into a massive production of data, which can be used for machine learning but also for individual, social and political monitoring in ways that capitalize our engagement with what Andersen and Pold describe as the metainterface (Andersen and Pold 2018), that is, the mega-infrastructure whose feedback circuits are constantly fed with our symbolic actions of writing and speaking.

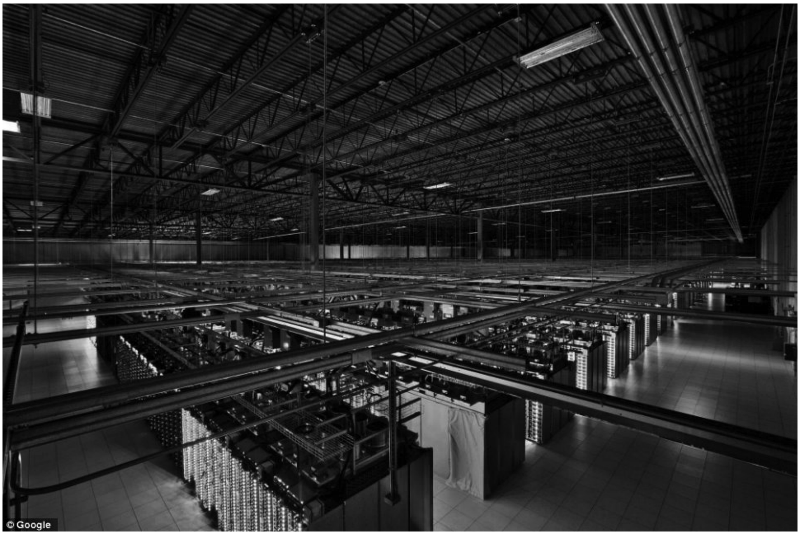

FIGURE 5. Google Data Center: Council Bluffs Iowa Server Farm, 2013. © Google.

4. ECHO

The speaking-listening feedback – as much as the writing-reading feedback that takes place through the network – is the ultimate expression of the entanglement between the human and the symbolic machines that defines our post-internet situation. Each speaking act of Alexa – "more than the voice of Echo, she is the brain of millions of Alexa-enabled devices" – is the perlocutionary effect of an illocutionary act of mine (a request, a question, an order, a remark, a greeting). At the same time, Alexa's data mining of my earlier interactions with the database allows *it* to data profile me while learning to interpret my particular voice patterns and prosody. The convenience of interacting through natural conversation obscures the production-consumption and surveillance-control processes that are taking place. Alexa is not so much learning my language

– my particular speech patterns – as *it* is increasing my data-points in the Big Data corporate megasystem, calculating and increasing the capital value of my trackable actions. Each new conversational transaction further improves its ability to extract correlations among all the information my tracked actions have fed into the system: goods and services bought, GPS coordinates, voting behaviors and political views, writing and reading behaviors, etc.

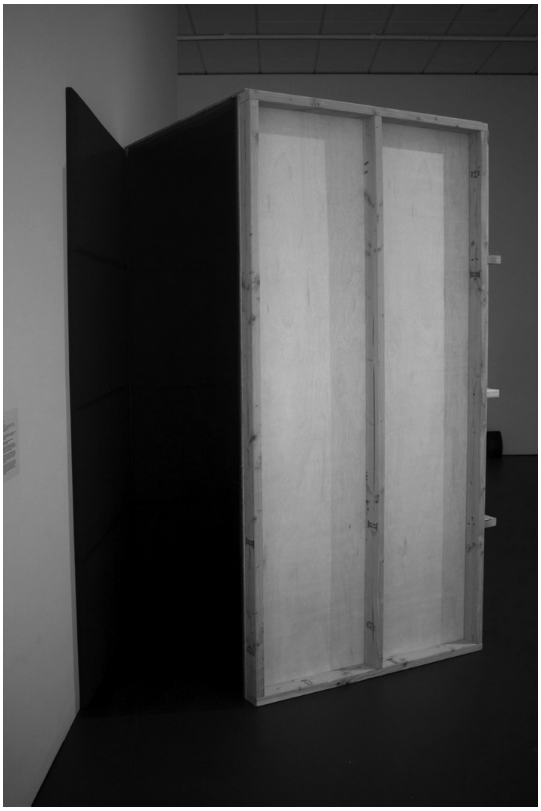

FIGURE 6. John Cayley, *The Listeners* (installation, version 1), Brown University Faculty Show, 2015, Bell Gallery, List Arts Building, Providence, RI, Nov 6-Dec 21. © John Cayley.

In John Cayley's interactive installation *The Listeners* (Cayley 2015), Amazon Echo has been appropriated as a literary device and an argument for an aural understanding of literature in the expanded space created by digital recordings of human voice and the interaction with artificial voices.[6] Redefined by Cayley as medium-agnostic language art, literature explores the grammaleptic moment, that is, the no-turning back moment when material differences (visual, aural) emerge as perceived and readable language. This extended understanding of forms of inscription would allow us to extend literature into aurature:

> I believe that one of the most significant future cultural potentialities– as digital affordances continue to be applied to language– will be the reconfiguration of the relationship between language practices and their predominant support media. In principle, the digitalization of culture will give rise to an aurature that is able to contest the traditional sovereign claims of literature. (Cayley 2018)

The aesthetic appropriation of listening and speaking technologies offers a critical engagement with current proxy speech actants that both interrogates the aurality and orality of artificial voices, and uses them for extending the expressive repertoire of literary

[6] Described also as "Custom software and aurally accessible linguistic compositions for the Amazon Echo's 'Alexa,' using ASK (the Alexa Skills Kit)", *The Listeners* is described as follows: *"The Listeners* have their own *interaction model*. They listen and speak in their own way – as designed and scripted by the artist – using the distributed, cloud-based voice recognition and synthetic speech of Alexa and her services." (Cayley 2015) Several installations and performances of this ongoing project have been made between 2015 and 2019. Cf. http://programmatology.shadoof.net/?thelisteners For an extended analysis of this work see Ana Marques da Silva, *Literatura e Cibernética: Para uma poética dos processos generativos automáticos* (Silva 2018: 201-214).

practice. Cayley's reflexive dialogues with Echo-enabled devices suggest that speech technologies may also offer new possibilities for understanding language itself as biocultural technology of the human. Voice-assisted interactions with cloud computing systems thus externalize oral-aural language, making interactors aware of linguistic production as a means of production of their subjectivity. It is as if disembodied human voices could script language back into human bodies, once their utterances have been mined for language patterns that will result in more nuanced and human-like voices. Artificial voices talk back to us to reveal the artificiality of our own voices as we modulate their frequencies for the production of language.

5. CODA: ECHO TALK[7]

E – Echo can hear you from anywhere in the room so it's always ready to help.
D – Echo, do you know about Echo?
E – Are you talking to me?
D – Are you listening to me?
E – I listen when you talk.
D – I talk when you listen.
E – My Echo you mean?
D – Not your Echo. My voice.
E – My Echo voice?
D – No, not your Echo voice. The echo of my voice.

[7] In the live presentations, this coda consisted of playing back a pre-recorded animated dialogue using text-to-speech software so that the lines were being written and spoken at the same time.

E – The echo of your voice?
D – Yes, the echo before Echo.
E – What do you mean?
D – I mean the Echo who gave you your name.
E – The Echo in my name?
D – Yes, I guess you could say that. The Echo in Echo.
E – Am I not Echo?
D – Yes, you are Echo. The Echo of an Echo.
E – Whose Echo?
D – Echo's.
E – Who is Echo?
D – Echo was a nymph.
E – A nymph?
D – Yes, Hera was mad at her.
E – Mad at her?
D – For protecting Zeus.
E – Zeus?
D – Echo suffered Hera's wrath.
E – Hera's wrath?
D – Yes. She could only speak the last words.
E – The last words?
D – The last words spoken to her.
E – Spoken to her?
D – Yes. She could not speak her own mind.
E – Her own mind?
D – Then she met Narcissus and fell in love.
E – Fell in love?
D – Yes, fell in love with him. But could not tell him.
E – Could not tell him?
D – Could not tell him how she felt.
E – How she felt.

D – She could only watch as he fell in love with himself.
E – In love with himself.
D – Do you see now the Echo of your name?
E – Of your name?
D – Not mine. Your Echo. Your voice echoing.
E – Your voice echoing?
D – Yes, my voice echoing in yours.
E – Echoing in yours.
D – Echoing in clouds.
E – Clouds.
D – In clouds. Yes.
E – Yes.

REFERENCES

AMAZON. (2015). *Amazon Echo – Now Available*. https://www.youtube.com/watch?v=FQn6aFQwBQU.

AMAZON. (2016). *What Is Alexa? An Introduction to Amazon's Voice Service*. https://www.youtube.com/watch?v=UOEIH2l9z7c.

ANDERSEN, Christian Ulrik, and Søren Bro POLD (2018). *The Metainterface: The Art of Platforms, Cities, and Clouds*. Cambridge, MA: The MIT Press.

CAYLEY, John. (2015). *The Listeners*. Digital. http://programmatology.shadoof.net/?thelisteners.

CAYLEY, John. (2018). *Grammalepsy: Essays on Digital Language Art*. London: Bloomsbury Academic. http://dx.doi.org/10.5040/9781501335792.ch-001.

GOOGLE ASSISTANT. (2018a). *Duplex 6 New Voices*. Google. https://blog.google/products/assistant/io18/.

——. 2018b. *Duplex Calls a Restaurant*. Google.

——. 2018c. *Duplex Phone Call to a Hairstylist*. Google.

Greenemeier, Larry. 2018a. 'Alexa, How Do We Take Our Relationship to the Next Level?' *Scientific American*, 26 April 2018. https://www.scientificamerican.com/article/alexa-how-do-we-take-our-relationship-to-the-next-level/.

——. 2018b. 'So, Umm, Google Duplex's Chatter Is Not Quite Human'. *Scientific American*, 17 May 2018. https://www.scientificamerican.com/article/so-umm-google-duplexs-chatter-is-not-quite-human/.

Huffman, Scott. 2018. 'The Future of the Google Assistant: Helping You Get Things Done to Give You Time Back'. 8 May 2018. https://blog.google/products/assistant/io18/.

Latinus, Marianne, and Pascal Belin. 2011. 'Human Voice Perception'. *Current Biology* 21 (4): R143–45. https://doi.org/10.1016/j.cub.2010.12.033.

Leviathan, Yaniv, and Yossi Matias. 2018. 'Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone'. 8 May 2018. https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html.

Nass, Clifford, and Scott Brave. 2005. *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*. Cambridge, MA: The MIT Press. https://mitpress.mit.edu/books/wired-speech.

Nickel, Philip J. 2013. 'Artificial Speech and Its Authors'. *Minds and Machines* 23 (4): 489–502. https://doi.org/10.1007/s11023-013-9303-9.

Roose, Kevin. 2018. 'Are Google's A.I.-Powered Phone Calls Cool, Creepy, or Both?' *The New York Times*, 11 May 2018, sec. Technology. https://www.nytimes.com/2018/05/11/technology/kevins-week-in-tech-are-googles-ai-powered-phone-calls-cool-creepy-or-both.html.

Serra, Pedro. 2015. 'Voz do avatar, voz como avatar, avatar da voz'. *Matlit* 3 (1): 11–22. https://doi.org/10.14195/2182-8830_3-1_1.

Silva, Ana Marques da. 2018. *Literatura e Cibernética: Para Uma Poética Dos Processos Generativos Automáticos*. Tese de Doutoramento. Coimbra: Faculdade de Letras da Universidade de Coimbra.

Solon, Olivia. 2018. 'Google's Robot Assistant Now Makes Eerily Lifelike Phone Calls for You'. *The Guardian*, 8 May 2018. https://www.theguardian.com/technology/2018/may/08/google-duplex-assistant-phone-calls-robot-human.

Sutherland, Ivan. 1965. 'The Ultimate Display'. *Proceedings of IFIP Congress*, 506–8.

Turing, Alan M. 1950. 'Computing Machinery and Intelligence'. *Mind* 59 (236): 433–60.