Transformations issue 29 (2017)
www.transformationsjournal.org

ISSN 1444-3775

The Materiality of the Digital and the Gendered Voice of Siri

Thao Phan

AUTHOR BIO

Thao Phan is a PhD candidate in the Media and Communications program at the University of Melbourne. Her research looks at gender and artificial intelligence. She is part of the Transformative Technologies Research Unit at the University of Melbourne, and was the Editor of the *Platform* special issue "A Manifesto for cyborgs thirty years on: Gender, technology and feminist-technoscience in the twenty-first century."

ABSTRACT

We live, contends Alexander Galloway, in an algorithmic culture. Algorithms are now inescapably embedded into everyday life transforming processes and objects from cultural artefacts into "smart" systems. But unlike most algorithms, which are obscured behind the black box of post-industrial processes, intelligent personal assistant softwares such as Apple's Siri are imbued with *voice* and *personality*. That is, they are given a materiality and tangibility. This paper aims to interrogate the nature of this materiality, and specifically, the manifestation of the gendered voice. It is my contention that the gendered voice of Siri is symptomatic of the difficulties in performing trust and transparency in what is essentially an intangible process. As Christian Sandvig has argued, transparency and trust are processes that must be seen in order to be believed but the issue with algorithms is that for the most part they *can't be seen*. Thus for these "robots," the performance of human sociality, specifically the use of language, humour, and the presentation of gender are cunning manoeuvres that contribute to the performance of trust in the theatre of persuasion. Continuing Sandvig's trajectory, this research seeks to explore the relationship between gender, sociality, and immediacy in these artificial systems.

KEYWORDS

Siri, gender, voice, algorithmic culture

Algorithms are often occluded from discussions of social robotics. Where social robots are associated with the embodied practices of nuanced social being in a material environment, algorithms are instead equated with the abstract and immaterial mechanics of digital infrastructure. They operate invisibly, generating search results, navigating databases, mapping preferences, making suggestions and curating newsfeeds. They are fundamental to the smooth functioning of search engines, social networking sites, recommender systems, banking processes, news and current affair feeds, and burgeoning techno-cultures such as gaming. In a prescient analysis, Alexander Galloway has described video games as "algorithmic cultural objects" (Galloway 6) flagging the ways in which algorithmic processes are embedded within the structures and frameworks of gaming techniques, practices and cultures. Indeed, the growing ubiquity of algorithms, in not just gaming but almost all everyday processes, has led to the questioning of algorithmic integrity. As Tarletan Gillespie argues, "Algorithms play an increasingly important role in selecting what information is considered most relevant to us, a crucial feature of our participation in public life" (Gillespie 167). Yet despite this critical role, there is still surprisingly little attention paid to algorithms in discussions of robotic sociality. What is a robot if not an algorithm machine? How can we think about human sociality within the context of algorithmic cultural objects? It is my contention that the failure to recognise algorithms within these discussions is in part due to the continuing tensions between the material and the digital. Where robots are thought of as tangible, material, and embodied, algorithms are associated with the abstract, virtual and disembodied.

The responsibility placed on algorithms to mediate public life has led to conscious campaigns to portray them as neutral scientific objects, divested of specialist interest. The languages of computationalism and scientific rationalism are co-opted in order to legitimate algorithmic authority. Attempts to eliminate human bias have translated to a superficial removal of human bodies from algorithmic discourse. The aim is, ultimately, to instantiate a sense of scientific "neutrality" through the displacement of human subjectivity with disembodied objectivity. Algorithms in this discourse represent the principles of automation and rationalism, indeed, they are the perfect conflation of the two, the literalisation of automated rationality. The very idea of automation invokes an anxiety around the displacement of bodies. Whether on the factory floor or in the decision-making process, the human element is said to be removed from the chain of command, and with it, the biases of the body. In short, without the human it is assumed that there can be no human partiality within an algorithmic process. This myth of neutrality conveniently serves the interests of those who produce the algorithms themselves. As Gillespie states, "the careful articulation of an algorithm as impartial (even when that characterisation is more obfuscation than explanation) certifies it as a reliable sociotechnical actor, lends its results relevance and credibility, and maintains the provider's apparent neutrality in the face of the millions of evaluations it makes" (Gillespie 179). This paper seeks to contribute to the literature on social robotics by opening the conversation to include algorithms as social

machines. I will interrogate Apple's intelligent personal assistant software Siri as a kind of "robot" that negotiates human sociality through gendered performances. Unlike most algorithms, which are essentially intangible actors, Siri is imbued with *voice* and *personality*, that is, it is given a materiality and tangibility. How does this materiality expose embedded biases regarding digital objects? In particular, how can the manifestation of the gendered voice be used to critique the assumed disembodied discourses of the digital? By analysing the gendered voice as a symptom of the desire to manufacture an invisible social actor, I seek to critique the assumed neutrality of algorithms. Indeed, by submitting a digital object to a politics of materiality, in this case Judith Butler's critique of gendered materiality, I wish to expose the inherent contradictions in separations between the digital and the material.

Neutrality and algorithms

Part of the difficulty in articulating algorithms as embodied social actors arises from the commercial characterisation of algorithms as disconnected from human intent. As authors such as Christian Sandvig and Tarleton Gillespie have argued, the centrality of algorithms in public life has translated to desperate efforts to communicate a sense of algorithmic objectivity and transparency. However, these characterisations are, in many ways, a willful deception. In his essay analysing the promotional rhetoric of popular algorithms, Sandvig critiques the visualisation tactics deployed by commercial organisations to manage their algorithm's "public image." Commenting on commercial discourse, he argues that "computer algorithms now have their own public relations. That is, they have a public-facing identity and new promotional discourses that depict them as efficient, valuable, powerful, and objective" (Sandvig). Attempts to communicate the complicated internal functions of algorithms to lay audiences has resulted in a rhetoric that obscures, defers and misleads as much as it illuminates, illustrates or reveals.

For instance, in their efforts to explain the process of their sorting algorithm, Google has employed a series of visual metaphors. The sorting algorithm is represented as an assembly line in which web pages are assessed on a rolling conveyor belt, queries are bashed into place using an industrial press, pages are given ticks of approval for quality by a robotic arm, and content is checked for "freshness" via a giant sniffing nose. Sandvig notes that the assembly line as dominant metaphor is an interesting irony given the computer industry's embrace of the myth of the post-industrial economy; one that displaces the manufacturing processes of Taylorism in favour of more flexible regimes of production. He also notes two major omissions from this authorised narrative: 1) the presence of advertisements and the role of sponsored content, and 2) the presence of human workers within the query process. In regards to the first omission, advertising content accounts for 90-95% of the Google revenue model (Sandvig). The willful distancing of this content from the depiction of the process can be understood as part of a

broader project to maintain the public integrity of the algorithm, and coextensively the organisation, as outside the domain of vested interest. Great care is taken to posit search outputs as unbiased, as matter-of-fact, and as the result of a scientifically proven procedure. The articulation of this procedure as precise and efficient through naming outputs as "best results" or "top stories" is closely tied to the corporate desire to manufacture an artificial sense of trustworthiness and transparency.

In regards to the second omission, I would argue that the removal of embodied humans within the assembly line process is a figurative gesture. In order to minimise anxiety concerning data security and privacy, the "impartial" algorithm here displaces the "partial" and biased eyes of the surveillant human. The careful curation of these visual metaphors to draw from some bodily apparatuses (arms and noses) but not others (eyes and ears) can be read as an effort (conscious or otherwise) to navigate the role of the body in embodied performances of trust. Where arms and noses connote operations of sensing and doing, eyes and ears are linked to more insidious manifestations of listening and watching – anxieties that loom large within the broader context of pervasive regimes of state and corporate surveillance, which companies such as Google are complicit within. These visual metaphors fracture a coherent sense of a materiality of algorithms by promoting a series of abstract metaphors that are not tied to any specific experience of embodiment.

Gillespie has similarly commented on the difficulties of defining algorithms themselves by critiquing definitional language. He argues that the term "algorithm" is misleading as it suggests a singular and stable object which one can hold to account in terms of impartiality; "what we might refer to as an algorithm is often not one algorithm but many" (Gillespie 178). Using the example of Google, he cites the common practice of A/B testing by which multiple iterations of the algorithm may be simultaneously released and then dynamically assessed and improved through mass feedback systems. As is the case with most web 2.0 applications, the idea of a finished and stable digital object is a delusive ideal exploited to give the impression of stability and security. Again, this slipperiness and commitment to neutrality contributes to a sense of abstractness that evades conventional ideas of embodied sociality. Indeed, in his critique, Gillespie argues that the very idea of a neutral algorithm is itself a "carefully crafted fiction" (Gillespie 179). Algorithms function at the behest of a variety of actors – commercial, legal, and political. In the case of Google, the performance of openness is ironically at odds with the company's highly protectionist and evasive behaviour. As Frank Pasquale notes, the details of the Google algorithm is a fiercely guarded trade secret and for that reason is not open to public inspection, however, to even postulate openness and accountability is here a deceit because the algorithm is "likely so dynamic that a snapshot of them would give us little chance of assessing their biases" (Pasquale).

For Sandvig and Gillespie, these knowing performances are acts of public legitimation. They function to allay anxiety over the already ubiquitous role

of algorithms in our daily lives by propagating an assurance conditional on the myth of neutrality. For them, the tension is between the claim to objectivity and the performance of objectivity itself. It is a dialectic that posits a perfectly neutral algorithm whose processes can, and should be, transparently communicated. However, it is my contention that neutrality itself is a myth, one that fallaciously configures bias as an externality that impinges on the neutral subject rather than as always already a part of any material process.

Gender is often taken as an example of bias. Particularly in discussions of technology, the invocation of the category of "gender-neutral" posits technological actors as *tabula rasa* – a clean slate or empty surface – as yet untouched by the biases of gender. The process of "gendering" technologies implies that gender is an externality, inscribed onto the neutral surface. Apple's Siri is often described in these terms. Gender, and especially the gendered voice, is only ever considered a supplementary or aesthetic aspect of the interface – an identity that is *a posteriori* decided upon and can also *a posteriori* be easily changed. Against this tenor, I argue that gender is an inherent part of understanding Siri as a material actor, an inherent part of the successful functioning of the interface itself and not a bias that is laid across a supposedly empty surface.

Siri and the gendered voice

In an algorithm such as Apple's Siri, the gendered voice is a prominent feature in the user interface. An integrated part of Apple's operating system, Siri ostensibly performs the functions of a personal assistant taking enquiries dictated by the user that range from questions about the weather, recommendations for restaurants and films, and requests for action such as setting alarms or reminders. As with any algorithm, Siri is not a stable actor but has gone through, and continues to go through, several iterations. The Siri program was acquired by Apple in 2007, and by 2011 was released as an integral feature of the iOS5 operating system on Apple iPhones. Since 2012, it has been a standard feature of all Apple mobile hardware, accessible across multiple platforms including mobile, tablet, Apple Watch, and Apple TV (Faber 2). Siri is voiced by a number of different voice actors whose accents are regional specific. The female voice was the default for most regions in its initial release (with the exception of Britain and France) and continues to be the dominant option in most regions (Faber 2). Despite the introduction of a "gender option" in iOS7 by which users could designate a male or female voice for Siri, the female voice remains the most common culturally associated voice of the program and is still the preferred voice used in promotional material for Apple.

As an algorithm, Siri is unusual because the gender politics implicit within the system are here openly negotiated. As I have so far argued, algorithms are often represented as abstract and immaterial. Siri, however, is not afforded the ambiguity of abstraction in the same way as, say, the Google sorting

algorithm. As an essential part of the interface, Siri *must* embody a voice and thus *must* confront the politics of a materiality. Spoken language falls under the larger umbrella term of natural user interfaces, a term that takes as its dialectic definition of the "natural" a sense of the innate and already present functions of human action. For example, gestures are classified as a natural user interface because they do not require the user to learn an artificial control system – gesture is considered an already apparent skill of the user. Similarly, spoken language is taken as natural because it demands little conditioning and so novice users can easily adapt and learn the system. A successful natural user interface is, then, one that does not challenge the user; it is one that is invisible to them. Invisibility here means diverting attention away from the act of mediation – it is the illusion of immediacy by which the subject is perfectly seduced by the medium, and in this seduction, indulges in the fantasy that there is no medium at all. As Bolter and Grusin have argued, "immediacy is transparency … it is the notion that a medium could erase itself and leave the viewer in the presence of the objects represented, so that he [sic] could know the objects directly" (Bolter and Grusin 70). In spoken language interfaces, there is an implicit appeal to the presumed immediacy of dialogue. It takes for granted that face-to-face communication is somehow more transparent. Part of the project of Siri is, then, the desire to be transparent – a manoeuvre that is complicit (or at least complements) the myth of neutrality and openness in algorithms. But there is an inherent tension in this desire. Whereas the natural user interface grasps at immediacy by displacing the medium onto the body (the body itself becomes the controller), Siri demands its own embodiment, its own voice.

So how is this tension between materiality and the invisible settled? Implicit in the successful functioning of a spoken language interface is the believable performance of human sociality. The category of the invisible becomes, then, a performance of the socially invisible. It is the perfect mimesis of the social order within the speech acts of the algorithm itself. We may ask, what is at stake in this category of the socially invisible? This is a question that is asked again and again in the studies of identity politics, by real bodies who are made invisible in social structures by virtue of the policing of race, gender, class, ability, sexuality and other markers of difference. In both instances, the tension is resolved through the same mechanism – by performing the category of the socially invisible. Whereas in the latter it is a strategy of survival – of passing or playing it straight or biting the tongue – in the former it is the result of willful desire to be transparent and invisible. Although I am by no means suggesting that invisibility is a desired category for a real body, there are instances in which the alternative to invisibility is an aggressive or unwelcome acknowledgement; a feeling of vulnerability and nakedness. Being congenial in these instances is a matter of preservation in the face of exhaustion. For the spoken language interface, however, performing the socially invisible means inverting these politics so that the socially invisible subject all of a sudden becomes *the most* desirable category.

The politics of materiality

In *Bodies that Matter*, Judith Butler discusses the politics of another kind of materiality; the materialisation of gendered and sexed bodies through the reiterative and performative practices of discursive regulatory norms. For Butler, materiality is inherently tied to power, as she states:

> What constitutes the fixity of the body, its contours, its movements, will be fully material, but materiality will be rethought as the effect of power, as power's most productive effect.… "Sex" is, thus, not simply what one has, or a static description of what one is: it will be one of the norms by which the "one" becomes viable at all, that which qualifies a body for life within the domain of cultural intelligibility. (Butler xii)

She argues against an understanding of performativity or construction as a position of choice (the positing of a choosing subject), but rather, construction as an effect of constitutive constraint that produces the domain of intelligibility by which some bodies are understood as more legitimate than others. It is not the subject who decides on gender; "on the contrary," she says, "gender is part of what decides the subject" (Butler ix). It is my contention, that Siri's subjectivication is entangled within these same processes. In its performance of the socially invisible, Siri requires gender to materialise itself as a subject.

Like all kinds of gendered materialisations, Siri is the result of a negotiation between the regulatory ideal and the body that it controls. Essential to this production of the intelligible subject is the simultaneous production of "abject beings" (Butler xiii). These are the subjects who are disavowed from discourse and through exclusionary means are rendered "not yet subjects." The subject is fragile and unstable and requires the constant positing of this *un*-subject, over and against which it can affirm itself as legitimate. For Siri, this is continuous confrontation with that which it cannot be, with moments of hyper-visibility by which the spell of immediacy is broken. In the context of the natural user interface and the algorithm, this means moments of distraction or illegibility.

In *Wired for Speech* Clifford Nass and Scott Brave explore the psychology of human voice perception and the implications for natural user interfaces. They remark on the common "social rules of voice" (3), that is, the normative standards that must be adhered to in order to effectively perform the category of social invisibility. Defying these rules draws attention to the interface rather than the output. If the user is distracted by the voice of the interface, they are less able to process the message itself. This is why regional accents and dialects are introduced into Siri's vernacular so that a British user would not be distracted by the program's use of Australian English. For a company like Apple, success in the interface is about reducing the number of these "distracting" encounters. Siri takes seriously the social rules of voice and for every instantiation of the algorithm there is the production of any

number of *un*-subjects that have been erased to refine and adhere to these rules.

Analysing the gendered voice of Siri, then, is a matter of looking not only at the nuances that are present, but the nuances that are also omitted. What is most visibly present is the domination of the female gendered voice in the Siri interface. On the one hand, this gendering can be understood as an obvious appeal to stereotypes regarding women in industry; a mimesis of gendered patterns of labour in which the embodiment of these intelligent personal assistant softwares gestures towards the real bodies of personal assistants, a profession still dominated by women. According to the Apple website (2016):

> Siri makes everyday tasks less taxing. It figures out which apps to use for which requests, and it finds answers to queries through sources like Yelp and Wolfram Alpha. It plays the songs you want to hear, gives you directions and wakes you up. All you have to do is ask.

Here, Siri is posited as a subservient and compliant subject. One who has no will of its own and who only performs the will of the user. It is an indictment on the domain of cultural intelligibility that the female gendered voice emerges as the "natural" choice for this role. If gender subjectivates the subject, then there is an inherent assumption regarding the social role of women that can interpellate the user as the one whose will is dominant within this dyad relation. On the other hand, we can shift the focus away from the female gendered voice to look at the presence of gender more broadly. The introduction of the "gender option" for Siri in 2013 meant that users could effectively change the gender of Siri's voice. The options for gender reflect the dominant Western binary system: male or female. So does the male-gendered voice instantiate the same position of social invisibility as the female-gendered? Not necessarily. Operating systems in which the default voice is female requires the user to consciously seek out and change this setting. This is a knowing act – a willful distancing of the self from the subject of the interface. Only the user who is aware of the "medium" – that is, someone who has not been seduced by immediacy – changes these default settings. In this case, there can be no appeal to the socially invisible subject as the category of invisibility itself is eschewed by the user.

It is worth noting that it makes no difference whether Siri itself identifies as having a gender. Indeed, if you ask the program its gender, it insists on its status as "genderless".

> Q:
> What gender are you?
>
> Siri:
> I am Siri.

> I don't have a gender.
> I am genderless. Like cacti. And certain species of fish.
> I was not assigned a gender.
> Animals and French nouns have genders. I do not.
> Don't let my voice fool you: I don't have a gender.
> I am still just … Siri.

What is signifiant, however, is the gendered relation it produces between itself and the user. As Butler notes, materiality is an effect of power, and the gendered materiality of the algorithm is wholly dependent on the power position of the user.

Breaking immediacy

In the above example, we can witness the ways in which Siri is programmed to give variance in its responses to indefinite or ambiguous questions. Some of these responses are abrupt, some are evasive, and some are even humorous. There is, however, a rhythm to these responses. Where possible, they avoid repetition within the syntax of each reply. Although this may seem like a banal observation, there has been a strong emphasis on behalf of the Apple organisation to carefully craft each variation and response. Jerome R. Bellegarda, senior human language technologies researcher at Apple, has discussed how these programmed language games contribute to a sense of anthropomorphism within the machine:

> One important aspect of the Siri experience is the system's response to user requests that fall outside of the well-defined domains it knows about. A typical example is a query such as "Siri, do you love me?", for which a factual response is besides the point. When faced with such out-of-domain queries, most systems typically exhibit a rather "clinical" behaviour, with responses like: "Sorry, I don't understand what you mean." While technically adequate, such behaviour lacks *humanness*. In contrast, Siri tries to provide somewhat more entertaining and/or whimsical responses.… Note that giving the same input three times in a row results in three different answers, as the same response would likely be annoying and otherwise destroy the illusion of anthropomorphism. (Bellegarda 11: emphasis added)

In what ways is this a more palatable experience? If we return to the question of immediacy, we can understand these attempts to produce "humanness" as also an attempt to produce the socially invisible subject. The algorithm is here mimicking the cadences of the human subject. To deliver what Bellegarda describes as repeated "clinical" responses would expose Siri not as a socially invisible subject but as a technological actor. To have for every answer "Sorry, I don't understand what you mean" repeated again and again would appear robotic; that is, inhuman. Unlike the Google algorithm, which elides the presence of human bodies in its discourse, in Siri we see the return of the human body into the domain of the digital.

This return is marked also by the politics of gender. The appeal to anthropomorphism is mediated by the successful performance of gender. Gender itself becomes a signifier for the body, for nature and the natural. Here it is the successful deployment of gender that coalesces a "natural" user interface. But there is a final irony here as gender, when taken as a sign of nature in this way, is exposed as anything other than "natural." Gender becomes a performance – indeed, it is *performative* – such that an algorithm can be programmed to follow its script. The tension between nature and culture here is all the more complicated by the terms of immediacy. For Bolter and Grusin, the logic of transparent immediacy operates on the desire to "diminish and ultimately deny the mediating presence" (23). Although this desire has manifested in different ways, it can never be wholly achieved. It is impossible to totally erase the mediation process itself. As with Siri, the artificial presence of the female-gendered voice has been repeatedly remarked on by users, such to the extent that to quiz Siri on questions of gender or sexual preference or other markers of an authentic *anthropos* has become a game, the results of which are shared and circulated online.

To conclude, what becomes increasingly clear in analysing Siri as a social actor is that gender comes to embody the algorithm. In this sense, there is no possibility of a materiality for this social robot that is not predicated on its subjectivisation through the gendered rules of sociality. Where in the popular discourse of algorithms, the body is elided in favour of the disembodied discourse of the virtual and the abstract, in this critique of Siri, the body itself is returned to the centre of discourse as an always already part of structuring the algorithm as a culturally intelligible subject. Algorithms are in some ways the ideal subject for the interrogation of social robots because they expose the contradictory nature of digital/material distinctions. Although they may not literally inhabit a stable body, their tenability is nevertheless contingent on body politics. In the case of Siri, it is the politics of gendered materiality that works to undermine and expose the ways in which robotic sociality appeals to the fallacy of immediacy.

Works Cited

Apple Inc. "iOS - Siri - Apple (AU)." Web. 14 Aug. 2016.

Bellegarda, Jerome R. "Spoken Language Understanding for Natural Interaction: The Siri Experience." *Natural Interaction with Robots, Knowbots and Smartphones*. Ed. Joseph Mariani et al. New York: Springer, 2014. 3–14.

Bolter, Jay David, and Richard Grusin. *Remediation: Understanding New Media*. Revised ed. edition. Cambridge, Mass.: The MIT Press, 2000.

Butler, Judith. *Bodies That Matter*. New York: Routledge, 2011.

Faber, Liz. *From Star Trek to Siri: (Dis)Embodied Gender and the Acousmatic Computer in Science Fiction Film and Television*. 2013. PhD thesis, Southern Illinois University Carbondale.

Galloway, Alexander. *Gaming*: *Essays on Algorithmic Culture.* Minneapolis and London: University of Minnesota Press, 2006.

Gillespie, Tarleton. "The Relevance of Algorithms." *Media Technologies: Essays on Communication, Materiality, and Society*. Ed. Tarleton Gillespie, Pablo J. Boczkowski, and Kirsten A. Foot. Cambridge, Massachusetts: The MIT Press, 2014. 167–194.

Pasquale, Frank. "Assessing Algorithmic Authority." *Madisonian*. 18 Nov (2009). < http://madisonian.net/2009/11/18/assessing-algorithmic-authority/>

Nass, Clifford, and Scott Brave. *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*. Cambridge, Mass: The MIT Press, 2005. Print.

Sandvig, Christian. "Seeing the Sort: The Aesthetic and Industrial Defense of "The Algorithm." *Media-N* 10.3 (2014). http://median.newmediacaucus.org/art-infrastructures-information/seeing-the-sort-the-aesthetic-and-industrial-defense-of-the-algorithm/.