

Reuniting Speech-Impaired People with Their Voices: Sound Technologies for Disability and Why They Matter for Organisation Studies

DOMENICO NAPOLITANO

domenico.napolitano-ssm@unina.it

Scuola Superiore Meridionale, Naples (Italy)

Abstract

This paper intends to provide an analysis of sound and voice technologies for speech-impaired people regarded as sites of knowledge production about disability. The study will focus on the case of Google's project to reunite speech-impaired users with their voices using voice cloning technology, an evolution of speech synthesis that allows the reconstruction of the sonic and timbral characteristics of a person's voice. Addressing both the narratives and representations – which reveal a medical model of disability as an external flaw to be cured through technology – and the material practices and operations enacted by those technologies – which highlight epistemologies of human variation, embodiment and accessibility built into the software –, the paper shows that disability as a social construct is co-constituted in the interaction of these levels. In this regard, the following research proposes a socio-technical model of disability theorisation that combines techno-scientific knowledge, cultural values, images of the user, material operations and organisational practices. From this perspective, the paper argues that the study of disabilities would benefit from the contribution of organisation studies and media studies in order to reveal the 'constructedness' of disability and able-bodiedness, and the role of media technologies, institutions, and representations in producing and upholding – as well as potentially challenging – such constructions.

1. Introduction

Released in December 2019, the documentary *The Age of A.I.* featured a Google project in voice technologies aiming “to reunite speech-impaired users with their original voice”.¹ The tech company worked with former NFL player Tim Shaw, who has lost his ability to speak after being affected by amyotrophic lateral sclerosis (ALS), a disease in which the neurons that control a person's voluntary muscles die, eventually leading to a total loss of control over one's body. As underlined by the documentary commenter, the project's aim was “to get his voice back”. Collecting voice recordings from Shaw's interviews on national television, Google employed Artificial Intelligence (specifically, *machine learning* techniques) to create a synthetic

¹ The statement can be found at the following webpage <https://deepmind.com/blog/article/Using-WaveNet-technology-to-reunite-speech-impaired-users-with-their-original-voices> (last accessed: June 6, 2021). The documentary is available online at https://www.youtube.com/watch?v=V5aZjsWM2wo&ab_channel=YouTubeOriginals (last accessed: July 20, 2021).

voice profile that imitates Shaw's way of speaking in a realistic and natural-sounding way (Chen *et al.*, 2019). Through a technology called *voice cloning* or, more generally, *voice synthesis*, Shaw could type sentences into a computer and have them expressed out loud with a synthesised version of his former voice, that is, with the sound of his voice before the disease "took it away from him". In the documentary, viewers can see the former football player, surrounded by his parents and three Google researchers, listening to the vocalisation of a letter he wrote to his younger self. It is a moment of intense emotion, underlined by Shaw's words, expressed with what remains of his own, natural voice: "[i]t has been so long since I've sounded like that. I feel like a missing part was put back in place".

Speech synthesis is already ubiquitous in the media landscape, employed in GPS navigation systems, Interactive Voice Responses (IVRs) in customer service hotlines and, more recently, in virtual assistants such as Siri or Alexa; but Stephen Hawking has also employed it since he lost the ability to speak in the 1980s due to ASL. The peculiarity of *voice cloning* is that it does not sound 'generic' as it did with Hawking's synthetic voice, aiming instead to reproduce the features of an embodied and personal voice, the voice of a specific person. This subtle but crucial passage is indicative of the complex connection between voice technologies and disability: such technologies pursue 'naturalness' and 'realism' to increase their usability and assistive power but, in so doing, they are also setting a specific way of treating disability, both in their way of imagining/representing 'users' and meeting their needs, and in their way of defining what a 'normal' voice is, how it should sound and what it should allow its 'users' to do.

In this paper, the case of voice technologies for speech-impaired people will be first examined, thus investigating sound media as sites of knowledge production about disability. Sound technology is a prominent component of social, cultural, and political change with the potential to allow broader accessibility and greater inclusion for disabled people with voice or hearing-related impairments. Nevertheless, it can also lead to new forms of exclusion or unintentional marginalisation, especially if sound technology is uncritically judged as a source of liberation without taking into consideration certain disabling features at the social and productive level. The following study will argue that this way of considering sound technologies, which can be extended to media in general, is theoretically productive for the implementation of a social model of disability (Oliver, 1996; Siebers, 2008; Goggin, 2009) and for challenging the long-standing use and adherence to the mainstream medical model. Nevertheless, the approach proposed in this paper further aims to underline the need for social models to focus on non-discursive and material practices. Such practices constitute the disabling constructed environment populated by media technologies – sharing instances arising from the materialist turn in social sciences, promoted in the fields of Science, Technology and Society (Woolgar, 1991), Actor-Network Theory (Law and Hassard, 1999), feminist material semiotics (Haraway, 1991) and media archaeology (Parikka, 2012). In this regard, the paper calls for collaboration between media studies and organisation studies in the framework of disability studies. In enabling communication and interaction for impaired people, voice and sound technologies are central for building *access* and *agency* and, therefore, transcend their status as media to become organisational structures. The discursive and material ways these technologies imagine, represent and empower their users and the organising practices they lead to thus producing socio-technical definitions of disability that affect cultural, political and institutional positions towards disability itself. I propose to call this attitude a *socio-technical model* of disability.

In the following sections, both the representations of disability, which emerge from the narrations of technology (e.g., the above-mentioned documentary), and the socio-material aspects of technological artefacts (e.g., the data-driven practices and operations employed in sound technologies for speech-impaired people) will be analysed. The investigation will highlight the intertwining of cultural and social aspects about ability and disability embedded in presentations, representations and discourses about voice technology with those built into algorithms, standards and material practices. This study will show how the medical modelling of the body, the design of media technologies and the agency built into organisational structures all shape one another and co-produce the socio-technical construct of disability.

Section 2 frames the contemporary debate about disability in organisation studies and media studies, with a particular focus on sound studies, highlighting possible convergences between the fields. Section 3, in particular, uses the case study of the cloning of Tim Shaw's voice by Google to reflect upon representations of disability in voice technology and AI, underlining the misleading promises of a purely technological fix typical of medical rhetoric. Section 4 discusses material practices and operations related to voice cloning in light of the metaphor of 'giving voice', which is so often used in mainstream narrations of disability as synonymous with giving 'access' and 'agency'. Finally, the Conclusion section of this paper is an invitation to read the main findings of the analysis as aspects of a socio-technical model of disability, a framework that calls for enhanced collaboration between organisation studies and sound and media studies.

2. The framework: disability, sound studies and organisation studies

Disability studies as a social and political movement interested in the lives and knowledge of disabled people as a marginalised group have gained significant recognition in the last decades (Barnes, 2004; Siebers, 2008; Ellcessor, Hagood and Kirkpatrick, 2017). Central to this field of studies is the critique of the medical model of disability, which considers it as "an individual defect lodged in the person, a defect that must be cured or eliminated if the person is to achieve full capacity as a human being" (Siebers, 2008: 3). Unlike the medical approach, which is individual, rehabilitative and productive (Williams and Mavin, 2012), the emerging field of disability studies considers disability as a *social construct*, in this way denaturalising it and promoting a social model of disability which does not aim to cure or avoid disease, but rather to study the social meanings, symbols, and stigmas attached to the concept of disability as a component of identity, and how it relates to enforced systems of exclusion and oppression.² According to this view, the concept of disability is socially created and is distinct from that of impairment, which describes the physical attribute. To quote Oliver's (1996: 22) famous claim:

² The use of the term 'disabled people' instead of 'people with disabilities' is consistent with the social model and, as such, is adopted in this paper. As well explained by Hamraie (2017: 280), "[d]isabled people' is the preferred usage for many disability activists, who argue that they are disabled by society and claim disability as an identity. 'Person-first language', such as 'people with disabilities', is often associated with medical and rehabilitative models, which identify personhood as separate from disabled embodiment and experience".

[...] we define impairment as lacking part or all of a limb, or having a defective limb, organ or mechanism of the body; and disability as the disadvantage or restriction of activity caused by a contemporary social organisation which takes little or no account of people who have physical impairments and thus excludes them from participation in the mainstream of social activities.

More recently, Tobin Siebers (2008: 3) has implemented this definition, arguing that disability is not an individual defect but “the product of social injustice, one that requires not the cure or elimination of the defective person but significant changes in the social and built environment”. This emphasis on the social and built environment is at the basis of the dialogue between disability studies and disciplines such as media studies and organisation studies in recent years. Suppose disability is constructed through normative assumptions underpinning socially constructed categories of difference. In that case, those assumptions are not located only in ideas and discourses but also built into material artefacts, technologies and organisational structures. Williams and Mavin (2012) argue that, although under-theorised in organisation studies, disability is an inherently organisational problem, not only because it solicits the adoption of solutions in workplaces and human resources management, but also because it is (Williams and Mavin, 2012: 160):

[...] hierarchically constructed, reproduced and maintained through organizing processes and practices in ways which center the requirements of some organizational members within normative expectations, while marginalizing the requirements of others.

Thus, the meeting of disability studies and organisation studies, especially in their critical declination, becomes an opportunity to explore how assumptions of non-disability (or ableism) infuse organising processes and are maintained as an organising norm, while organising is presumed to be neutral (Mumby, 2008). This means focusing on socio-material processes of organising which include norms, codes, legislations and knowledge, but also architectures, environments, artefacts and technologies (Czarniawska, 2015). As Hamraie (2017) argues, design-knowledge is of central importance since it reproduces assumptions of able-bodiedness, resulting in the marginalisation of disabled people. Furthermore, knowledge is inscribed in the design of objects and tools which can provide or deny *access* for certain people, thus becoming inextricably interwoven with operative power. On the other hand, ‘access-knowledge’ is a theoretical approach focused on examining the relations of knowledge and power inscribed into the design of objects, architectures and infrastructures that articulate and organise access and agency for people.

Placing artefacts and technologies at the centre of issues of disability, this approach to the study of disability has interested scholars in media studies as well (Goggin and Newell, 2003; Ellis and Kent, 2011; Ellcessor, Hagood and Kirkpatrick, 2017). Disability is all too often reconstituted and reconfigured in and through new technologies, both at the narrative and the operational level. As Ellis and Kent (2011) argue, new technologies can broaden accessibility and inclusion for impaired people through their affordances. However, those same affordances can also produce new forms of exclusion and disability. In this regard, Ellcessor, Hagood and Kirkpatrick (2017: 16) argue that “we need to understand the ways that media and information technologies are intertwined with the standardization and regulation of the human body”, and how those processes shape the meaning of ability and disability. In

Ellcessor, Hagood, and Kirkpatrick's (2017) view, this investigation must consider both meaning and materiality since ideas are embodied in and shaped by material conditions and human practice, and are given meaning by the discourses that inform them and that they in turn inform. Thus, innovations in media technologies raise issues of materiality and embodiment that also affect cultural ideas and meanings: how we interact with buttons, dials, or gaming consoles; how we plug in earbuds or position ourselves toward screens, listening devices or speaking devices; how manufacturers imagine the bodies that will engage with their creations, and so on.

These are also organisational issues since they concern the way bodies share physical and social space and access services and structures, how they communicate and act, but also how they are regulated through standards built into norms and artefacts. As Brunsson, Rasche and Seidl (2012) argue, standardisation is a form of organising outside formal organisations since standards are 'translated' into concrete practices, which include the design, adoption and use of media technologies. These practices, in turn, can be disabling for certain people and reproduce ableist marginalisation within organising. As noted by Mills and Sterne (2017: 365, emphasis in the original), "not only do media produce disability through their textual representations of disability, they produce disability through their very operations, their institutional existences, and their policy and juridical dimensions". Therefore, we must take great care with both representations and material operations, ideas and embodied practices, collective images and technical standards. In this sense, technology, as a "crystallization of socially organized action" (Sterne, 2003b: 367), can be thought of "as a 'strategic research site' for studying society and the organization of social practice. It is not only technology but also the culture of technology that can reproduce social inequality" (Alper, 2017: 11).

This approach is no less relevant when discussing sound technologies, as seen in the recent development of sound studies (Pinch and Bijsterveld, 2012; Sterne, 2012). For example, Mara Mills (2011, 2012, 2020) has recounted the shared history of sound devices such as the telephone and the audiometer, in which the instrumental measurement of 'normal' human hearing was developed to allow for the more efficient transmission of speech over telephone lines. Through her research, she demonstrates how the modern concepts of impairment and hearing loss, as well as the contributions of deaf and hard-of-hearing people, were central to the development of telecommunications technologies and signal processing in the 20th century. Similarly, Jonathan Sterne's first book, *The Audible Past* (2003a), locates the origins of sound reproduction in 19th-century sound culture, with its peculiar conceptions of hearing, speech, and deafness. In both cases, technologies and engineering had profound impacts on definitions of deafness and hearing impairment, as well as on definitions of what is a 'normal' voice. At the same time, users with disabilities have often found themselves at the forefront of innovations in media systems that make them more useful for everyone: "[i]deas of disability shaped the emergence of modern sound media and modern sound media shaped ideas of ability and disability" (Mills and Sterne, 2017: 369).

The focus on sound technologies helps us to identify specific dynamics of legibility and validation related to impairment and disability. Mack Hagood (2017) suggests that mediation provides a form of validation of acoustic impairments such as tinnitus which are invisible and, as such, hardly legible in social terms. According to Hagood (2017), our response to such conditions is exemplary of media's tendency to be at the same time a vector of social and collective identity construction, but also a vector of indirect medicalisation, functioning as a

space where impaired people can look for normalisation. This dynamic resembles that highlighted by Moser and Law (2003: 511):

First there is the desire of those who are able to 'normalise' those who are not. And then second, there is a loop: the desire by many of those who are disabled to be counted as competent or 'normal' by embracing normatively approved features of modern subjectivities.

Kafer (2013) noted that the rejection of the medical model in favour of the social one faces the risk of overlooking the lived experiences of impairment while politically marginalising disabled people interested in medical interventions or cures. In terming her model 'political/relational', Kafer (2013) highlights the need to attend disability as the result of dynamics of power that shape particular relationships among people, institutions, culture, and material structures such as media, technologies and medical practices.

Taking inspiration from these perspectives, this paper argues that the *social model* promoted by disability studies could be improved by expanding it and turning it into a *socio-technical model* that considers material and technical practices crucial parts of the social construction of disability. Since such practices are very often enacted within both the medical context (i.e., through techno-scientific medical instruments) and the organising context (i.e., through accessibility to certain structures or services, communication between people, groups and institutions, the definition of standards, etc.), this approach should carefully examine the processes of co-production of the disability construct in the interweaving of medical and scientific knowledge, technologically mediated agency, and social and institutional statements. By adopting this approach, disability studies would benefit from the contribution of organisation studies and media studies in order to reveal the constructedness of disability and able-bodiedness and the role of media technologies, institutions, and representations in producing and upholding – as well as potentially challenging – such constructions and ideologies.

In the following sections, this proposal will be articulated with regard to sound and voice technologies for disability, using as a case study Google's project of reuniting speech impaired people with their voices via voice cloning.

3. Representing disability through voice technology

In her taxonomy of representational strategies for depicting persons with disability, Rosemarie Garland Thomson (2001) delineates four visual rhetorical strategies in the representation of disability: the wondrous, the sentimental, the exotic, the realistic. The documentary featuring Google's project of reuniting speech-impaired people with their voices through voice technology can be ascribed to the *sentimental rhetoric*, addressing what Oliver (1996) defines as 'the tragedy model' of disability, which positions disability as a terrible issue for the afflicted individual. "Every single day is a struggle for me", Tim Shaw declares through his synthetic cloned voice. Disability is represented here as an inherently 'negative ontology' (Campbell, 2005). The emphasis emerging is that of an individualised medical interpretation of disability as a biological or functional limitation; the narration leads to an attempt to *compensate* people with impairments rather than *enabling* their full inclusion in social life.

Technology, particularly AI, is seen here as a tool to achieve this compensation and reduce the discursively constructed 'tragic' nature of disability, a tool to 'fix' the impaired person, thus overlooking the social conditions in which disability is represented as such. It is not by chance that the title of the documentary episode is *Healed through AI*: thanks to AI, 'a missing part' can be 'put back into place'. In this example, a specific way of understanding the 'cure' can also be noticed: as a victory over a malfunction of the body which, nonetheless, does not affect the inner core of personhood. Consequently, disability is considered an 'external' flaw or defect, and as such, it can be (and should be) cured. "Yes, my body is failing but my mind is not giving up", is what Shaw states in the documentary, using the cloned voice to speak.

Beyond the arguably noble intent of reuniting impaired people with their voices, careful deconstruction and analysis of Google's narration offers valuable insights into how we treat and understand disability in our society. Through such analysis, it is possible to unearth hidden connections between social constructions and cultural-epistemological attitudes that drive modern science and technology, of which AI is perhaps one of the most prominent outcomes. Google's narration seems indeed consistent with a particular type of AI ideology: that human beings might be 'uploaded' into new hardware whenever their old 'hardware' wears out (Natale and Pasulka, 2019). This deceitful claim (Natale, 2021) is grounded on a metaphysical understanding of reality insofar as it assumes the duality of mind and body, and the superiority of the former over the latter. This attitude denies both the social construction of disability and the embodied condition of subjectivity and identity, locating subjectivity instead in an abstract and supernatural 'soul'. Being disembodied, this kind of subjectivity is by definition 'able' and cannot be affected or defined by any impairment, which remains by definition an external, bodily flaw. Furthermore, this disembodied subjectivity can be transferred from one body to another. Voice, with its traditional relation to metaphysical subjectivity (Derrida, 2010 [1967]), is the perfect representation of this condition: the possibility to transfer voice from a biological body to an electronic one, as speech synthesis does (but also between different biological bodies through the computational medium, as it will be argued in the next section), is the metonymy of the supernatural status of inner subjectivity, of which voice is indeed the main expression. Voice has been considered by Western philosophy as the unmediated expression of an inner self, an attitude Derrida has named 'phonocentrism' or 'metaphysic of presence'. In this perspective, the loss of the voice would be the worst tragedy imaginable, a direct attack to the core of what makes us human. However, voice is also considered a bodily thing, a mere interface with the purity of the inner self, which is in fact disembodied. As such, the loss of voice does not compromise the supposed ability of the inner self. From this perspective, a compensation for the loss of the voice can be in the first place imagined and then pursued by medical or technological means.

A similar attitude transpires from the recent case of Hollywood actor Val Kilmer who, having lost his voice after a tracheotomy due to throat cancer, contracted with the company Sonantic to develop a clone of his former voice, using a similar technological process as Google's. "I'm still the same person, the same soul [...]. Now, with artificial voice, I can express myself again", the actor declared (Flynn, 2021).³ The self to which he refers is the inner, untouched self of metaphysics.

³ The statement is available online at https://www.youtube.com/watch?v=OSMue60Gg6s&ab_channel=Sonantic (last accessed: August 10, 2021).

This dualistic and metaphysical way of thinking requires the body to be considered only when separated from the 'soul'. The condition of uncorrupted ability granted on faith to that inner self is actually a myth, a 'supplement' (Derrida, 2010 [1967]) that must necessarily be located outside the real body in an imaginary body, an abstract self or a universal idea of ability. However, if the self is disconnected from the body and not articulated with it, disability becomes less severe because it can be considered neither essential nor constitutive of the subject itself. This reproduces an ideology of ability as "fear of disability" (Siebers, 2008: 9), which is typical of a medical and rehabilitative model where personhood is considered as separate and abstract from disabled embodiment and experience – and as such, not defined by it. Instead, taking embodiment seriously means overpassing a metaphysical (and ableist) perspective and facing disability as inherent to the human condition, a form of diversity that produces identity. As argued by Judith Butler (1990), and the many scholars interested in the intersection between gender, race and disability who have built on her work, it is not possible to make sense of the construction of the person unless the body also forms part of the picture. Such an intellectual attitude has political consequences as well.

There are, therefore, two aspects that emerge from Google's representation of disability. On the one hand, a technophile approach can be observed, which is informed by a kind of technological determinism that is classic of transhumanist ideologies: technology can, by itself, solve social problems, heal people, give access and opportunity. But, on the other hand, such an approach seems grounded in an "ideology of ability" (Siebers, 2008: 8) that assumes an original, untouched condition of ability and perfection, subsequently corrupted by disability, which must be restored through medical intervention and technology.

Such an approach assumes disability as an individual imperfection to be cured rather than a social construction to be confronted at the level of social justice. This model is rehabilitative and productive, aimed at fixing individuals to make them able again and ready to be (re)integrated into society, rather than questioning how the 'ableist' social context itself has (discursively) *disabled* them. In both Google's narration and Kilmer's words, the rhetoric of empowerment is viewed as an individualistic condition rather than a social one. It is the individual impaired person who is lacking and expected to conform to 'being normal', while the social context is regarded as unproblematic (Swain, French and Cameron, 2003).

According to this interpretation, Google and other tech companies seem to affirm a medical model through their representations, producing *narratives of personal liberation via technology* that overlook the social construction of disability. These stories portray technology as allowing individuals to 'overcome' their disability as an individual limitation and, at the same time, are intended to be uplifting and inspirational for able-bodied audiences, assuming ability as an unquestioned and neutral condition in the first place. For this reason, these narratives appear implicitly and essentially ableist. One consequence of this understanding of disability as an essentially individual problem is that the (marginalised) subject position of disabled people is not problematised. As Ingunn Moser (2006: 373) puts it:

[...] the mobilization of new technologies works to build an order of the normal and turn disabled people into competent normal subjects. However, this strategy based on compensation achieves its goals only at a very high price: by continuing to reproduce boundaries between abled and disabled, and normal and deviant, which constitute some people as disabled in the first place.

In her work about mobile technology and disability, Meryl Alper (2017: 23) notes that technological fixes intended to eliminate impairment presume a specific ‘image of the user’, which views disabled people as needing to be fixed. As underlined by several scholars (Siebers, 2008; Alper, 2017; Sterne, 2021), this condition reflects a paradox inherent to *assistive technology*. All technology, indeed, is assistive to a certain degree. Nevertheless, technology is defined as ‘assistive’ only when designed and marketed with impaired people in mind. In this way, ‘assistive technology’ assumes a specific and ideological representation of its intended users, who are implicitly defined as ‘people who need assistance’. This is a powerful example of how an ideology of ability *disables* impaired people through its discourses and representations, and of how the construct of disability works to marginalise certain people and groups. Could the growing popularity of ‘voice assistants’ such as Siri or Alexa be undermining this narration by blurring the boundaries between kinds of people who need assistance? In general, unmasking disability as a construct – that is, through the deconstruction of mainstream representations – serves to denaturalise dynamics of exclusion, undermining the assumption of ability as a normal condition while problematising the very concept of ability, especially in a technological landscape in which we are all assisted and ‘augmented’ by technology.

Of course, disability is not constructed only through representations. Indeed, the metaphor of ‘giving voice’ opens to other perspectives if analysed from the point of view of material practices.

4. ‘Giving voice’ between metaphor and materiality

In Western cultures, the term ‘voice’ is used both to symbolise human speech and as a powerful metaphor for agency, authenticity, truth, and self-representation (Hirschman, 1970). Alper (2017) noted that, in communication technologies, these two meanings often converge. Since the 1970s and the 1980s, technological devices that allowed speech-impaired people to communicate have been represented as ‘giving voice’ to those who cannot physically speak, subtly implying that technology is also metaphorically empowering those who have been silenced by normative and ableist environments. However, with voice technologies such as speech synthesis and voice cloning, the literal sense of ‘giving voice’ seems to surpass even the metaphorical one. In particular, voice cloning is presented as a technology that can give speech-impaired people their own voice back. Indeed, as already outlined, it does not produce merely a generic synthetic voice that allows people who have lost their voice to express words audibly; rather, it reconstructs the voice of a specific person, with the sonic, timbral and prosodic characteristics which make it unique and personal. Voice cloning, it could be argued, is like a mask used to customise and personalise already existing communication technologies based on speech synthesis, such as text-to-speech, voice interfaces and other assistive tools, in order to give them a touch of realism and naturalness (Napolitano, 2020a). In this way, such technology overcomes typical limits related to the cold and robotic features of synthetic voices, acquiring a ‘human touch’ which makes it more acceptable and usable, better able to deal with the four aspects of identity performance (as proposed by Alper, 2017: 51): (a) sounding one’s age; (b) sounding human; (c) sounding related to one’s family members; and (d) sounding consonant with one’s cultural background.

However, is this enough to ‘give voice’ to someone? Can a voice that sounds like mine be considered really mine? And, more in general, what does it mean to ‘own’ a voice?

Commodification is the name of the condition necessary for something to be ‘owned’, ‘given’ and ‘given back’. Therefore, a commodification of the voice is necessary in order to give it (back) to someone or to reunite someone with it. Recent studies (Connor, 2000; Sterne, 2003a; Ernst, 2016; Napolitano, 2020a) have investigated the social, cultural, historical and epistemological conditions of the commodification of the voice in different technological periods, analysing the different media through which voice has been stored, manipulated and re-enacted, from alphabetic writing to speaking automata, from the phonograph to the computer. At the very basis of these studies, there is the consideration that voice is never the pure immediate expression of an inner self, as stated by the metaphysics of presence, but it is always mediated. Although the idea of commodification seems to point towards ontology, the commodification of the voice is in fact a recognition that voice is not a static entity but, rather, something defined socio-technically by the intersection of media, culture and forms of embodiment (Moser and Law, 2003: 491; see also Napolitano, 2020b):

‘Voices’ do not exist in and of themselves. They do not reflect something that is pre-given. Rather they are constituted or ‘articulated’ into being in material arrangements which include social, technological and corporeal relations.

Therefore, to understand the specific socio-technical definition of voice enacted by voice cloning techniques, it is crucial to consider the assemblage of material practices and operations of the medium, together with discourses and metaphors. By adopting a constructionist perspective, a definition of voice-impairment can be derived directly from this assemblage since it is only within it that the idea of a ‘normal’ and a ‘personal’ voice acquires its sense. Therefore, it is only within the socio-technical assemblage that the sense of ‘giving voice’ can be obtained and understood so as to deconstruct it.

For the purpose of our discussion, it is important to highlight the following material aspects. Indeed, voice cloning employs machine learning algorithms (the most current declination of AI) to determine the features of a person’s voice profile – those characteristics which differentiate it from that of another person – starting from the *data* of that voice (Marr, 2018). This has two consequences:

1. The collection of such voice data becomes crucial to the successful operation of the technology. This leads to the articulation of socio-material practices of voice data collection, such as *voice banking*, which must be pre-emptive. That is, they must happen before someone experiences significant speech impairment (or as soon as one is diagnosed with a degenerative disease). This has implications that go beyond the simple act of speaking, associating voice banking with cultural practices related to saving, preventing, leaving traces, making a will (not so far from having to decide in advance whether to be an organ donor).
2. Voice cloning relies on a sort of essentialist epistemology based on the assumption that voice is something that can be *owned*. It assumes that each person owns a voice, their own voice, and this voice can be translated into data with little or no loss from the ‘original’, embodied one. This means affirming a quantitative correspondence between voice and sound data and between sound data and identity, all of which has practical

and political consequences such as the possibility to profile people from the measurement of their voice features, ever more included in forensic practices of automatic speaker identification (Singh, 2021).

Voice technologies, in general, and voice cloning, in particular, enact specific forms of embodiment:

1. A synthetic voice does not come from the person's mouth or body but rather from a loudspeaker. The assumption that a voice that does not come from someone's mouth can still be considered as someone's own voice is something that destabilises naturalised ideologies of 'presence' (Derrida, 2010 [1967]) and universalist conceptions of the body, the voice and the identity (Sterne, 2021). Those conceptions are grounded on a latent normative assumption about what a voice is and where it comes from, assumptions that can be reproduced and even enhanced by sound media. Speech synthesis algorithms, for instance, need to measure bodily features of voice and mouth and extract standards from those measures in order to artificially recreate someone's voice (Napolitano, 2020b). In so doing, sound technologies "render abstract human qualities as measurable quantities" (Mills and Sterne, 2017: 369), defining operatively and scientifically what a voice is and how it should sound.
2. At the same time, and in an equally relevant way, voice cloning technology seems to take a step beyond the normative perspective: as vocal features are made to emerge directly from the data, which is a prerogative of machine learning algorithms, this technology is not attempting to synthesise a 'pre-defined' standard voice, but rather to get to the singularity of each voice, in its peculiarities and without reference to standards, also capturing the 'imperfections' and deviations from the norm that make each voice unique. Although recalling imperfection is still a way to re-enact an ableist model, things like unsupervised learning and deep learning seem to rely less and less on standards, models and norms, moving more and more towards immanently emergent behaviours. While this opens to a wide debate about a possible end of theory (Anderson, 2008; Napolitano, 2020b) that goes beyond the scope of this paper, it is nonetheless relevant that the material operations by which such technologies function offer a glimpse of an epistemological shift. This makes AI look like a promising technology when discussing disability because, instead of looking for norms among the variations, it opens the technology to the differences that characterise disability as itself "a form of human variation" (Siebers, 2008: 25) – a material practice that collides with the mainstream deceptive narration about AI described in the previous section. One consequence of this can be observed in that voice cloning, bypassing classic issues with synthetic voice, sidesteps the problem of avoiding gender, race and class stereotypes culturally coded into pre-defined voice characters. The fact that features of a person's body can be built into a synthetic voice (indeed, these very factors make that voice personal) represents more than mere engineering of the human body with rehabilitative aims (Hamraie, 2017: 15): it can also be seen as another form of technological embodiment.
3. Despite this, voice cloning does not seem to question the idea of 'normal speaking ability' it reproduces through text-to-speech systems. Even if the *sound* of the voice is individual, the *way of speaking* performed by those systems is still relatively standardised, based on certain social expectations, which are in turn based on ideas of ability. This might be a purely technological gap, which will be soon compensated by

new features that allow users to control the way of speaking (e.g., Sonantic is already working on such software). But it should also be seen as another clue of how sound media shape ideas of ability and disability through their very operations, while disability, in turn, shapes ideas about sound media (Sterne and Mills, 2017). This co-production of media and disability is evident when considering that, while being employed to treat speech impairment, voice cloning is also spreading as a cultural practice in the media environment, from worrying *deepfake* phenomena in which the voice of someone is cloned to be used in frauds – that is, to make digital voices say things their human inspirations never uttered (Wilson, 2018) – to *voice skins* in which people, especially videogamers, can swap their voice in real-time with that of others (MODULATE, 2020). In all these examples, new forms of technological embodiments can be noticed that short-circuit with the traditional assumptions of able-bodiedness, above all, the logocentric one that assumes the natural and unmediated correlation of body and voice.

Therefore, while still characterised by a certain medical rhetoric, the spreading of voice cloning and the specific material operations and practices it enacts are shaping disability as a social construct insofar as they question assumptions of embodiment and ability even beyond the context of speech impairment.

In light of these considerations, ‘giving voice’ assumes a peculiar meaning. More than a metaphor, ‘giving voice’ becomes a social and cultural practice based on the material possibility of loosening the ableist model that considers voice, body and subjectivity as given and intrinsically bonded together. Voice technologies such as speech synthesis play a crucial part in this shift since they offer examples of the way in which voice is articulated into being in material arrangements that include social, technological and corporeal relations.

To complete the picture, the socio-economic conditions necessary so as to allow ‘giving voice’ as well as ‘giving back the voice’ should not be underestimated. Someone can ‘give back’ something only from a position of power, and in the case of AI-based voice technology, the power to give back the voice comes from a position of absolute techno-economic domination of tech giants like Google. To give back the voice means to think of the voice as something that can be possessed, but also as something that can be subtracted: rendering voice into data entails both being able to reconstruct it, seize it, recognise it and capture it into neoliberal dynamics of profiling and surveillance (Turow, 2021). Being in the position to ‘give back the voice’, these companies, in a certain sense, ‘dispose’ of the voice of those who have lost it; they appropriate it (mainly in the form of data). This corresponds, metaphorically but also materially, to the reproduction of a power relationship in which those who can ‘give back’ can also ‘take away’ and ‘reduce to silence’. As Moser and Law (2003: 499) underline, “‘giving a voice’ also takes away other possible voices”.

Based on all of these elements, technologies for disability seem characterised by a certain tension: while enhancing agency and access for impaired people, such technologies often reproduce a paradigm of rehabilitation that subdues individuals to a direct or indirect system of power based on normalisation and productivity. In this regard, Alper’s (2017: 2) words seem a good closure:

‘Giving voice to the voiceless’ regularly stands in for the idea that the historically disadvantaged, underrepresented, or vulnerable gain opportunities to organize, increase

their visibility, and express themselves by leveraging the affordances of information, media, and communication technologies [...]. These tools may selectively amplify voices within and across various publics and audiences, but their existence does not automatically call the status quo of structural inequality (i.e., racism, patriarchy, misogyny, and homophobia) into question.

5. Conclusion

In this paper, a detailed description has been provided of how sound media, and voice technologies in particular, shape ideas of disability while, in turn, being shaped by disability. This co-constitution takes place through and within discourses, representations and material practices of organising, programming, using and embodying voice through technology. Instead of considering these technologies just for their assistive potential, they have been regarded as places where knowledge about disability is produced, represented and enacted. In this sense, the parallel analysis of different aspects allows cultural and social ideas about disability embedded within them to emerge, highlighting how the persistence of a medical model of disability, especially in the representations of technology, is intertwined with social considerations emerging from material practices. On the one hand, narratives of personal liberation through technology, such as those proposed by Google or Sonantic, reveal the persistence of an ableist model grounded in a metaphysical idea of the bond between voice and subjectivity, resulting in a marginalisation of disability as an external flaw to be cured. On the other hand, practices of data collection and technological embodiment enacted by voice cloning reveal an epistemological sensitivity to human variation, leading to more inclusive understandings of disability.

In my perspective, these instances should not be considered as in opposition. Instead, I see them as co-constitutive of the social construction of disability. While this perspective is consistent with what has been defined as the social model of disability, I have proposed to call it a socio-technical model in order to underline its emphasis on the complex assemblage of knowledge, cultural values, collective images of the user, material operations and organisational practices built into technologies and through which disability is shaped and defined.

From such a perspective, the study of voice and sound technology can be meaningful both as a site of knowledge production about disability and as an example of the productive collaboration between media studies and organisation studies in the framework of disability studies. This is particularly evident since organisations are the sites where such knowledge is translated into individual and collective agency through policies and laws as well as through accessibility and communication practices and strategies.

Moreover, considering the centrality assumed by material processes in this perspective, such a collaboration could direct the theoretical and social analysis of the technologies of disability toward a better and more inclusive design of both artefacts and their organisational contexts, one that accommodates a wide range of bodies without reproducing dis/abled divides – one that keeps disabled people in mind.

Keywords:

sound technology; speech synthesis; voice cloning; disability studies; media studies; organisation studies

References

- Alper, M. (2017), *Giving Voice: Mobile Communication, Disability, and Inequality*, Cambridge, MA: MIT Press.
- Anderson, C. (2008, June 23), "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete", in *Wired*. Available online at http://www.wired.com/science/discoveries/magazine/16-07/pb_theory (last accessed: May 19, 2020).
- Barnes, C. (2004), "Disability, Disability Studies and the Academy", in J. Swain, S. French, C. Barnes, and C. Thomas (eds), *Disabling Barriers, Enabling Environments*, pp. 28–33, London: SAGE.
- Brunsson, N., Rasche, A., and Seidl, D. (2012), "The Dynamics of Standardization: Three Perspectives on Standards in Organization Studies", *Organization Studies*, 33 (5-6): 613–632.
- Butler, J. (1990), *Gender Trouble: Feminism and the Subversion of Identity*, New York/London: Routledge.
- Campbell, F.K. (2005), "Legislating Disability: Negative Ontologies and the Government of Legal Identities", in S. Tremain (ed.), *Foucault and the Government of Disability*, pp. 108–130, Ann Arbor, MI: The University of Michigan Press.
- Chen, Y., Casagrande, N., Zhang, Y., and Brenner, M., (2019), "Using WaveNet Technology to Reunite Speech-Impaired Users with Their Original Voices", *Google DeepMind* [weblog]. Available online at <https://deepmind.com/blog/article/Using-WaveNet-technology-to-reunite-speech-impaired-users-with-their-original-voices> (last accessed: June 30, 2021).
- Connor, S. (2000), *Dumbstruck: A Cultural History of Ventriloquism*, Oxford: Oxford University Press.
- Czarniawska, B. (2015), *A Theory of Organizing*, Cheltenham: Edward Elgar Publishing.
- Derrida, J. (2010 [1967]), *Voice and Phenomenon: Introduction to the Problem of the Sign in Husserl's Phenomenology*, trans. by L. Lawlor, Evanston, IL: Northwestern University Press.
- Ellcessor, E., Hagood, M., and Kirkpatrick, B. (2017), "Introduction: Toward a Disability Media Studies", in E. Ellcessor and B. Kirkpatrick (eds.), *Disability Media Studies*, pp. 1–28, New York, NY: New York University Press.
- Ellis, K., and Kent, M. (2011), *Disability and New Media*, New York, NY: Routledge.
- Ernst, W. (2016), *Sonic Time Machines: Explicit Sound, Sirenic Voices and Implicit Sonicity*, Amsterdam: Amsterdam University Press.
- Flynn, J. (2021), "Helping Actor Val Kilmer Reclaim His Voice", *Sonantic* [weblog]. Available online at <https://www.sonantic.io/blog/helping-actor-val-kilmer-reclaim-his-voice> (last accessed: August 10, 2021).

- Garland Thomson, R. (2001), "Seeing the Disabled: Visual Rhetorics of Disability in Popular Photography", in P.K. Longmore and L. Umansky (eds.), *The New Disability History: American Perspectives*, pp. 355–74, New York, NY: New York University Press.
- Goggin, G. (2009), "Disability and the Ethics of Listening", *Continuum: Journal of Media & Cultural Studies*, 23 (4): 489–502.
- Goggin, G., and Newell, C. (2003), *Digital Disability: The Social Construction of Disability in New Media*, Lanham, MD: Rowman & Littlefield.
- Hagood, M. (2017), "Disability and Biomediation: Tinnitus as Phantom Disability", in E. Ellcessor and B. Kirkpatrick (eds), *Disability Media Studies*, pp. 311–329, New York, NY: New York University Press.
- Hamraie, A. (2017), *Building Access: Universal Design and the Politics of Disability*, Minneapolis, MN: University of Minnesota Press.
- Haraway, D. (1991), *Simians, Cyborgs and Women: The Reinvention of Nature*, London: Free Association Books.
- Hirschman, A.O. (1970), *Exit Voice and Loyalty: Responses to Decline in Firms, Organizations, and States*, Cambridge, MA: Harvard University Press.
- Kafer, A. (2013), *Feminist, Queer, Crip*, Bloomington, IN: Indiana University Press.
- Law, J., and Hassard, J. (eds) (1999), *Actor Network Theory and After*, Oxford: Blackwell.
- Marr, B. (2019, May 6), "Artificial Intelligence Can Now Copy Your Voice: What Does That Mean for Humans?", *Forbes*. Available online at <https://www.forbes.com/sites/bernardmarr/2019/05/06/artificial-intelligence-can-now-copy-your-voice-what-does-that-mean-for-humans/> (last accessed: July 20, 2021).
- Mills, M. (2011), "Deafening: Noise and the Engineering of Communication in the Telephone System", *Grey Room*, 43: 118–143.
- Mills, M. (2012), "Do Signals Have Politics? Inscribing Abilities in Cochlear Implants", in T. Pinch and K. Bijsterveld (eds), *The Oxford Handbook of Sound Studies*, pp. 320–346, New York, NY: Oxford University Press.
- Mills, M. (2020), "Testing Hearing with Speech", in V. Tkaczyk, M. Mills and A. Hui (eds), *Testing Hearing: The Making of Modern Aurality*, pp. 23–48, New York, NY: Oxford University Press.
- Mills, M., and Sterne, J. (2017), "Afterword II: Dismediation – Three Proposals, Six Tactics", in E. Ellcessor and B. Kirkpatrick (eds), *Disability Media Studies*, pp. 365–378, New York, NY: New York University Press.
- Moser, I. (2006), "Disability and the Promises of Technology: Technology, Subjectivity and Embodiment within an Order of the Normal", *Information, Communication & Society*, 9 (3): 373–395.
- Moser, I., and Law, J. (2003), "'Making Voices': New Media Technologies, Disabilities, and Articulation", in G. Liestøl, A. Morrison, and T. Rasmussen (eds), *Digital Media Revisited*, pp. 491–520, Cambridge, MA: MIT Press.
- MODULATE (2020), "Voice Skins and Individual Identity", *Modulate* [weblog]. Available online at <https://www.modulate.ai/blog/voice-skins-and-individual-identity> (last accessed: July 25, 2021).

- Mumby, D.K. (2008), "Theorizing the Future of Critical Organization Studies", in D. Barry and H. Hansen (eds), *The SAGE Handbook of New Approaches in Management and Organization*, pp. 27–28, Los Angeles, CA: SAGE.
- Napolitano, D. (2020a), "The Cultural Origins of Voice Cloning", in M. Verdicchio, M. Carvalhais, L. Ribas and A. Rangel (eds), *xCoAx 2020 Proceedings of the Eighth Conference on Computation, Communication, Aesthetics & X*. Available online at <http://2020.xcoax.org/xcoax2020.pdf> (last accessed: July 30, 2021).
- Napolitano, D. (2020b), "'Where's the Voice of the Machine?' An Ethnography of Artificial Voice Socio-Technical Networks", *Etnografia e Ricerca Qualitativa*, 3: 351-372.
- Natale, S. (2021), *Deceitful Media: Artificial Intelligence and Social Life after the Turing Test*, New York, NY: Oxford University Press.
- Natale, S., and Pasulka, D. (2019), *Believing in Bits: Digital Media and the Supernatural*, New York, NY: Oxford University Press.
- Oliver, M. (1996), *Understanding Disability: From Theory to Practice*, Basingstoke: Macmillan.
- Parikka, J. (2012), *What is Media Archeology?*, Cambridge: Polity Press.
- Pinch, T., and Bijsterveld, K. (eds) (2012), *The Oxford Handbook of Sound Studies*, New York, NY: Oxford University Press.
- Siebers, T. (2008), *Disability Theory*, Ann Arbor, MI: University of Michigan Press.
- Singh, R. (2019), *Profiling Humans from Their Voice*, Singapore: Springer.
- Sterne, J. (2003a), *The Audible Past: Origins of Sound Reproduction*, Durham/London: Duke University Press.
- Sterne, J. (2003b), "Bourdieu, Technique, and Technology", *Cultural Studies*, 17 (3-4): 367–389.
- Sterne, J. (ed.) (2012), *The Sound Studies Reader*, New York, NY: Routledge.
- Sterne, J. (2021), *Diminished Faculties: A Political Phenomenology of Impairment*, Durham/London: Duke University Press.
- Swain, J., French, S., and Cameron, C. (2003), *Controversial Issues in a Disabling Society*, Buckingham: Open University Press.
- Turow, J. (2021), *Voice Catchers: How Marketers Listen in to Exploit Your Feelings, Your Privacy, and Your Wallet*, New Haven, CT: Yale University Press.
- Williams, J., and Mavin, S. (2012), "Disability as Constructed Difference: A Literature Review and Research Agenda for Management and Organization Studies", *International Journal of Management Reviews*, 14: 159–179.
- Wilson, M. (2018, June 3), "The War on What's Real", *The fast Company*. Available online at <https://www.fastcompany.com/90162494/the-war-on-whats-real> (last accessed: November 15, 2020).
- Woolgar, S. (1991), "The Turn to Technology in Social Studies of Science", *Science, Technology, and Human Values*, 16 (1): 20–50.